

10/074527

B1
#5

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12P 21/04, C12N 15/00, C07H 21/02	A1	(11) International Publication Number: WO 00/55351 (43) International Publication Date: 21 September 2000 (21.09.00)
(21) International Application Number: PCT/US00/05883 (22) International Filing Date: 8 March 2000 (08.03.00) (30) Priority Data: 60/124,270 12 March 1999 (12.03.99) US (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Laytonsville, MD 20882 (US). (74) Agents: WALES, Michele, M. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>
(54) Title: HUMAN COLON CANCER ASSOCIATED GENE SEQUENCES AND POLYPEPTIDES (57) Abstract <p>This invention relates to newly identified colon or colon cancer related polynucleotides and the polypeptides encoded by these polynucleotides herein collectively known as "colon cancer antigens", and to the complete gene sequences associated therewith and to the expression products thereof, as well as the use of such colon cancer antigens for detection, prevention and treatment of disorders of the colon, particularly the presence of colon cancer. This invention relates to the colon cancer antigens as well as vectors, host cells, antibodies directed to colon cancer antigens and recombinant and synthetic methods for producing the same. Also provided are diagnostic methods for diagnosing and treating, preventing and/or prognosing disorders related to the colon, including colon cancer, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of colon cancer antigens of the invention. The present invention further relates to methods and/or compositions for inhibiting the production and/or function of the polypeptides of the present invention.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Human Colon Cancer Associated Gene Sequences and Polypeptides

5

Field of the Invention

This invention relates to newly identified colon or colon cancer related polynucleotides and the polypeptides encoded by these polynucleotides herein collectively known as "colon cancer antigens," and to the complete gene sequences
10 associated therewith and to the expression products thereof, as well as the use of such colon cancer antigens for detection, prevention and treatment of disorders of the colon, particularly the presence of colon cancer. This invention relates to the colon cancer antigens as well as vectors, host cells, antibodies directed to colon cancer antigens and recombinant and synthetic methods for producing the same. Also
15 provided are diagnostic methods for diagnosing and treating, preventing and/or prognosing disorders related to the colon, including colon cancer, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of colon cancer antigens of the invention. The present invention further relates to methods and/or compositions for
20 inhibiting the production and/or function of the polypeptides of the present invention.

Background of the Invention

Colorectal cancers are among the most common cancers in men and women in the U.S. and are one of the leading causes of death. Other than surgical resection no
25 other systemic or adjuvant therapy is available. Vogelstein and colleagues have described the sequence of genetic events that appear to be associated with the multistep process of colon cancer development in humans (Trends Genet 9(4):138-41 (1993)). An understanding of the molecular genetics of carcinogenesis, however, has not led to preventative or therapeutic measures. It can be expected that advances in
30 molecular genetics will lead to better risk assessment and early diagnosis but colorectal cancers will remain a deadly disease for a majority of patients due to the

lack of an adjuvant therapy. Adjuvant or systemic treatments are likely to arise from a better understanding of the autocrine factors responsible for the continued proliferation of cancer cells.

Colorectal carcinoma is a malignant neoplastic disease. There is a high incidence of colorectal carcinoma in the Western world, particularly in the United States. Tumors of this type often metastasize through lymphatic and vascular channels. Many patients with colorectal carcinoma eventually die from this disease. In fact, it is estimated that 62,000 persons in the United States alone die of colorectal carcinoma annually.

At the present time the only systemic treatment available for colon cancer is chemotherapy. However, chemotherapy has not proven to be very effective for the treatment of colon cancers for several reasons, the most important of which is the fact that colon cancers express high levels of the MDR gene (that codes for multi-drug resistance gene products). The MDR gene products actively transport the toxic substances out of the cell before the chemotherapeutic agents can damage the DNA machinery of the cell. These toxic substances harm the normal cell populations more than they harm the colon cancer cells for the above reasons.

There is no effective systemic treatment for treating colon cancers other than surgically removing the cancers. In the case of several other cancers, including breast cancers, the knowledge of growth promoting factors (such as EGF, estradiol, IGF-11) that appear to be expressed or effect the growth of the cancer cells, has been translated for treatment purposes. But in the case of colon cancers this knowledge has not been applied and therefore the treatment outcome for colon cancers remains bleak.

There is a need, therefore, for identification and characterization of such factors that modulate activation and differentiation of colon cells, both normally and in disease states. In particular, there is a need to isolate and characterize additional molecules that mediate apoptosis, DNA repair, tumor-mediated angiogenesis, genetic imprinting, immune responses to tumors and tumor antigens and, among other things, that can play a role in detecting, preventing, ameliorating or correcting dysfunctions or diseases of the colon.

Summary of the Invention

The present invention includes isolated nucleic acid molecules comprising, or alternatively, consisting of, a colon and/or colon cancer associated polynucleotide sequence disclosed in the sequence listing (as SEQ ID Nos: 1 to 773) and/or contained
5 in a human cDNA clone described in Tables 1, 2 and 5 and deposited with the American Type Culture Collection ("ATCC"). Fragments, variant, and derivatives of these nucleic acid molecules are also encompassed by the invention. The present invention also includes isolated nucleic acid molecules comprising, or alternatively consisting of, a polynucleotide encoding a colon or colon cancer polypeptide. The
10 present invention further includes colon and/or colon cancer polypeptides encoded by these polynucleotides. Further provided for are amino acid sequences comprising, or alternatively consisting of, colon and/or colon cancer polypeptides as disclosed in the sequence listing (as SEQ ID Nos: 774 to 1546) and/or encoded by a human cDNA clone described in Tables 1, 2 and 5 and deposited with the ATCC. Antibodies that
15 bind these polypeptides are also encompassed by the invention. Polypeptide fragments, variants, and derivatives of these amino acid sequences are also encompassed by the invention, as are polynucleotides encoding these polypeptides and antibodies that bind these polypeptides. Also provided are diagnostic methods for diagnosing and treating, preventing, and/or prognosing disorders related to the colon,
20 including colon cancer, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of colon cancer antigens of the invention.

Detailed Description

25

Tables

Table 1 summarizes some of the colon cancer antigens encompassed by the invention (including contig sequences (SEQ ID NO:X) and the cDNA clone related to the contig sequence) and further summarizes certain characteristics of the colon
30 cancer polynucleotides and the polypeptides encoded thereby. The first column shows the "SEQ ID NO:" for each of the 773 colon cancer antigen polynucleotide sequences of the invention. The second column provides a unique "Sequence/Contig ID"

identification for each colon and/or colon cancer associated sequence. The third column, "Gene Name," and the fourth column, "Overlap," provide a putative identification of the gene based on the sequence similarity of its translation product to an amino acid sequence found in a publicly accessible gene database and the database accession no. for the database sequence having similarity, respectively. The fifth and sixth columns provide the location (nucleotide position nos. within the contig), "Start" and "End", in the polynucleotide sequence "SEQ ID NO:X" that delineate the preferred ORF shown in the sequence listing as SEQ ID NO:Y. The seventh and eighth columns provide the "% Identity" (percent identity) and "% Similarity" (percent similarity), respectively, observed between the aligned sequence segments of the translation product of SEQ ID NO:X and the database sequence. The ninth column provides a unique "Clone ID" for a cDNA clone related to each contig sequence.

Table 2 summarizes ATCC Deposits, Deposit dates, and ATCC designation numbers of deposits made with the ATCC in connection with the present application.

Table 3 indicates public ESTs, of which at least one, two, three, four, five, ten, fifteen or more of any one or more of these public EST sequences are optionally excluded from certain embodiments of the invention.

Table 4 lists residues comprising antigenic epitopes of antigenic epitope-bearing fragments present in most of the colon or colon cancer associated polynucleotides described in Table 1 as predicted by the inventors using the algorithm of Jameson and Wolf, (1988) Comp. Appl. Biosci. 4:181-186. The Jameson-Wolf antigenic analysis was performed using the computer program PROTEAN (Version 3.11 for the Power MacIntosh, DNASTAR, Inc., 1228 South Park Street Madison, WI). Colon and colon cancer associated polypeptides shown in Table 1 may possess one or more antigenic epitopes comprising residues described in Table 4. It will be appreciated that depending on the analytical criteria used to predict antigenic determinants, the exact address of the determinant may vary slightly. The residues and locations shown in Table 4 correspond to the amino acid sequences for most colon and colon cancer associated polypeptide sequence shown in the Sequence Listing.

Table 5 shows the cDNA libraries sequenced, and ATCC designation numbers and vector information relating to these cDNA libraries.

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

5 In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of
10 matter, or particular cell is not the original environment of the polynucleotide. The term "isolated" does not refer to genomic or cDNA libraries, whole cell total or mRNA preparations, genomic DNA preparations (including those separated by electrophoresis and transferred onto blots), sheared whole cell genomic DNA preparations or other compositions where the art demonstrates no distinguishing
15 features of the polynucleotide/sequences of the present invention.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X (as described in column 1 of Table 1) or the related cDNA clone (as described in column 9 of Table 1 and contained within a library deposited with the ATCC). For example, the polynucleotide can contain the
20 nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having an amino acid sequence encoded by a polynucleotide of the invention as broadly defined (obviously excluding poly-Phenylalanine or poly-
25 Lysine peptide sequences which result from translation of a polyA tail of a sequence corresponding to a cDNA).

In the present invention, "SEQ ID NO:X" was often generated by overlapping sequences contained in multiple clones (contig analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X is deposited at Human
30 Genome Sciences, Inc. (HGS) in a catalogued and archived library. As shown in column 9 of Table 1, each clone is identified by a cDNA Clone ID. Each Clone ID is unique to an individual clone and the Clone ID is all the information needed to

retrieve a given clone from the HGS library. In addition to the individual cDNA clone deposits, most of the cDNA libraries from which the clones were derived were deposited at the American Type Culture Collection (hereinafter "ATCC"). Table 5 provides a list of the deposited cDNA libraries. One can use the Clone ID to
5 determine the library source by reference to Tables 2 and 5. Table 5 lists the deposited cDNA libraries by name and links each library to an ATCC Deposit. Library names contain four characters, for example, "HTWE." The name of a cDNA clone ("Clone ID") isolated from that library begins with the same four characters, for example "HTWEP07". As mentioned below, Table 1 correlates the Clone ID names
10 with SEQ ID NOs. Thus, starting with a SEQ ID NO, one can use Tables 1, 2 and 5 to determine the corresponding Clone ID, from which library it came and in which ATCC deposit the library is contained. Furthermore, it is possible to retrieve a given cDNA clone from the source library by techniques known in the art and described elsewhere herein. The ATCC is located at 10801 University Boulevard, Manassas,
15 Virginia 20110-2209, USA. The ATCC deposits were made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for the purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to
20 sequences contained in SEQ ID NO:X, or the complement thereof (e.g., the complement of any one, two, three, four, or more of the polynucleotide fragments described herein), and/or sequences contained in the related cDNA clone within a library deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42 degree C in a solution comprising 50% formamide, 5x SSC
25 (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65 degree C.

Also included within "polynucleotides" of the present invention are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower
30 stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt

conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37 degree C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4 ; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50 degree C with
5 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress
10 background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

15 Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically
20 any double-stranded cDNA clone generated using oligo dT as a primer).

The polynucleotides of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-
25 stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A
30 polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of

modifications can be made to DNA and RNA: thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

In specific embodiments, the polynucleotides of the invention are at least 15, at least 30, at least 50, at least 100, at least 125, at least 500, or at least 1000 continuous nucleotides but are less than or equal to 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb, 7.5kb, 5 kb, 2.5 kb, 2.0 kb, or 1 kb, in length. In a further embodiment, polynucleotides of the invention comprise a portion of the coding sequences, as disclosed herein, but do not comprise all or a portion of any intron. In another embodiment, the polynucleotides comprising coding sequences do not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene of interest in the genome). In other embodiments, the polynucleotides of the invention do not contain the coding sequence of more than 1000, 500, 250, 100, 50, 25, 20, 15, 10, 5, 4, 3, 2, or 1 genomic flanking gene(s).

"SEQ ID NO:X" refers to a colon cancer antigen polynucleotide sequence described in Table 1. SEQ ID NO:X is identified by an integer specified in column 1 of Table 1. The polypeptide sequence SEQ ID NO:Y is a translated open reading frame (ORF) encoded by polynucleotide SEQ ID NO:X. There are 773 colon cancer antigen polynucleotide sequences described in Table 1 and shown in the sequence listing (SEQ ID NO:1 through SEQ ID NO:773). Likewise there are 773 polypeptide sequences shown in the sequence listing, one polypeptide sequence for each of the polynucleotide sequences (SEQ ID NO:774 through SEQ ID NO:1546). The polynucleotide sequences are shown in the sequence listing immediately followed by all of the polypeptide sequences. Thus, a polypeptide sequence corresponding to polynucleotide sequence SEQ ID NO:1 is the first polypeptide sequence shown in the sequence listing. The second polypeptide sequence corresponds to the polynucleotide sequence shown as SEQ ID NO:2, and so on. In other words, since there are 773 polynucleotide sequences, for any polynucleotide sequence SEQ ID NO:X, a corresponding polypeptide SEQ ID NO:Y can be determined by the formula $X + 773 = Y$. In addition, any of the unique "Sequence/Contig ID" defined in column two of Table 1, can be linked to the corresponding polypeptide SEQ ID NO:Y by reference to Table 4.

The polypeptides of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

The colon and colon cancer polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced

polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

5 It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The colon and colon cancer polypeptides of the present invention are
10 preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified using techniques described herein or otherwise known in the art, such as, for example, by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from
15 natural, synthetic or recombinant sources using techniques described herein or otherwise known in the art, such as, for example, antibodies of the invention raised against the polypeptides of the present invention in methods which are well known in the art.

By a polypeptide demonstrating a "functional activity" is meant, a polypeptide
20 capable of displaying one or more known functional activities associated with a full-length (complete) protein of the invention. Such functional activities include, but are not limited to, biological activity, antigenicity [ability to bind (or compete with a polypeptide for binding) to an anti-polypeptide antibody], immunogenicity (ability to generate antibody which binds to a specific polypeptide of the invention), ability to
25 form multimers with polypeptides of the invention, and ability to bind to a receptor or ligand for a polypeptide.

"A polypeptide having functional activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular assay, such as,
30 for example, a biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to

the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention).

5 The functional activity of the colon cancer antigen polypeptides, and fragments, variants derivatives, and analogs thereof, can be assayed by various methods.

For example, in one embodiment where one is assaying for the ability to bind or compete with full-length polypeptide of the present invention for binding to an
10 antibody to the full length polypeptide antibody, various immunoassays known in the art can be used, including but not limited to, competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, in situ immunoassays
15 (using colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another
20 embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

In another embodiment, where a ligand is identified, or the ability of a
25 polypeptide fragment, variant or derivative of the invention to multimerize is being evaluated, binding can be assayed, e.g., by means well-known in the art, such as, for example, reducing and non-reducing gel chromatography, protein affinity chromatography, and affinity blotting. See generally, Phizicky, E., et al., Microbiol. Rev. 59:94-123 (1995). In another embodiment, physiological correlates polypeptide
30 of the present invention binding to its substrates (signal transduction) can be assayed.

In addition, assays described herein (see Examples) and otherwise known in the art may routinely be applied to measure the ability of polypeptides of the present

invention and fragments, variants derivatives and analogs thereof to elicit polypeptide related biological activity (either in vitro or in vivo). Other methods will be known to the skilled artisan and are within the scope of the invention.

5

Colon and Colon Cancer Associated Polynucleotides and Polypeptides of the Invention

It has been discovered herein that the polynucleotides described in Table 1 are expressed at significantly enhanced levels in human colon and/or colon cancer tissues.

- 10 Accordingly, such polynucleotides, polypeptides encoded by such polynucleotides, and antibodies specific for such polypeptides find use in the prediction, diagnosis, prevention and treatment of colon related disorders, including colon cancer as more fully described below.

- 15 Table 1 summarizes some of the polynucleotides encompassed by the invention (including contig sequences (SEQ ID NO:X) and the related cDNA clones) and further summarizes certain characteristics of these colon and/or colon cancer associated polynucleotides and the polypeptides encoded thereby.

Table 1

Seq ID No.	Sequence/ Contig ID	Gene Name	Overlap	HGS Nucleotide Start End	% Identity	% Similarity	Clone ID
1	500802	Immunoglobulin kappa light chain variable region L25 [Homo sapiens] >pir S41816 S41816 Ig kappa chain V region L25 - human Length = 119 (AB008790) Grb7V protein [Homo sapiens] >sp D103000 D103000 GRB7V PROTEIN. >gil 526535 Grb7 protein [Homo sapiens] {SUB 130-343} Length = 447	gil415381	2 304	73	86	HGBA183
2	531091			2 292			HUKDY21
3	553147			3 440			HCASG85
4	558860	(AF039700) antigen NY-CO-38 [Homo sapiens] >sp G317020 G317020 ANTIGEN NY-CO-38. >gil 3170198 (AF039699) antigen NY-CO-37 [Homo sapiens] {SUB 1-403} Length = 652 MDA-7 [Homo sapiens] >sp Q13007 MDA7_HUMAN MDA-7 PROTEIN PRECURSOR (MELANOMA DIFFERENTIATION ASSOCIATED PROTEIN 7). Length = 206 disintegrin-protease [Homo sapiens] >sp O15204 O15204 DISINTEGRIN- PROTEASE. Length = 470	gillPID d1030000	33 635	97	98	HCEGY28
5	561730	(AF039700) antigen NY-CO-38 [Homo sapiens] >sp G317020 G317020 ANTIGEN NY-CO-38. >gil 3170198 (AF039699) antigen NY-CO-37 [Homo sapiens] {SUB 1-403} Length = 652 MDA-7 [Homo sapiens] >sp Q13007 MDA7_HUMAN MDA-7 PROTEIN PRECURSOR (MELANOMA DIFFERENTIATION ASSOCIATED PROTEIN 7). Length = 206 disintegrin-protease [Homo sapiens] >sp O15204 O15204 DISINTEGRIN- PROTEASE. Length = 470	gil3170200	34 393	98	98	HSDFA48
6	585938	(AF039700) antigen NY-CO-38 [Homo sapiens] >sp G317020 G317020 ANTIGEN NY-CO-38. >gil 3170198 (AF039699) antigen NY-CO-37 [Homo sapiens] {SUB 1-403} Length = 652 MDA-7 [Homo sapiens] >sp Q13007 MDA7_HUMAN MDA-7 PROTEIN PRECURSOR (MELANOMA DIFFERENTIATION ASSOCIATED PROTEIN 7). Length = 206 disintegrin-protease [Homo sapiens] >sp O15204 O15204 DISINTEGRIN- PROTEASE. Length = 470	gil1141751	206 538	81	81	HMQBR31
7	587785	(AF039700) antigen NY-CO-38 [Homo sapiens] >sp G317020 G317020 ANTIGEN NY-CO-38. >gil 3170198 (AF039699) antigen NY-CO-37 [Homo sapiens] {SUB 1-403} Length = 652 MDA-7 [Homo sapiens] >sp Q13007 MDA7_HUMAN MDA-7 PROTEIN PRECURSOR (MELANOMA DIFFERENTIATION ASSOCIATED PROTEIN 7). Length = 206 disintegrin-protease [Homo sapiens] >sp O15204 O15204 DISINTEGRIN- PROTEASE. Length = 470	gillPID c332729	2 331	100	100	HOSBO86

211	836789	GP36b glycoprotein [Homo sapiens] >pir G01447 G01447 GP36b glycoprotein - human >sp Q12907 Q12907 GP36B GLYCOPROTEIN PRECURSOR. Length = 356	gij505652	1	849	99	99	99	HKAAD74
212	838577	binding protein [Oryctolagus cuniculus] >gij182628 FK506-binding protein (FKBP) [Homo sapiens] >gij182633 FKBP-12 protein [Homo sapiens] >gij182649 FK506-binding protein 12 [Homo sapiens] >gij288196 FKBP [Homo sapiens] >gij665650 FK-506 binding protein [H	gij165023	2	433	100	100	100	HCRQD09
213	838717			676	900				11E8UJ03
214	839008			2	997				HF0XSS2
215	840063	(AF006751) ES/130 [Homo sapiens] >sp O75300 O75300 ES/130. Length = 977	gij3299885	3	2729	84	85	85	HWL1IX68
216	840533			183	482				HWLLU74
217	840669			474	1115				HPMGM71
218	841140	(AF081281) lysophospholipase [Homo sapiens] >sp O75608 O75608 LYSOPHOSPHOLIPASE. Length = 230 polypeptide GalNAc transferase-T4 [Mus musculus] >sp O08832 O08832 POLYPEPTIDE GALNAc TRANSFERASE-T4. Length = 578	gij3415123	1	789	100	100	100	HAIJCC51
219	841386		gij2121220	491	1258	66	81	81	HMCCA66
220	841480			3	212				HDQET68
221	841509			3	662				HTELO87
222	841616			340	660				HWLFT95

The first column of Table 1 shows the "SEQ ID NO:" for each of the 773 colon cancer antigen polynucleotide sequences of the invention.

The second column in Table 1, provides a unique "Sequence/Contig ID" identification for each colon and/or colon cancer associated sequence. The third column in Table 1, "Gene Name," provides a putative identification of the gene based on the sequence similarity of its translation product to an amino acid sequence found in a publicly accessible gene database, such as GenBank (NCBI). The great majority of the cDNA sequences reported in Table 1 are unrelated to any sequences previously described in the literature. The fourth column, in Table 1, "Overlap," provides the database accession no. for the database sequence having similarity. The fifth and sixth columns in Table 1 provide the location (nucleotide position nos. within the contig), "Start" and "End", in the polynucleotide sequence "SEQ ID NO:X" that delineate the preferred ORF shown in the sequence listing as SEQ ID NO:Y. In one embodiment, the invention provides a protein comprising, or alternatively consisting of, a polypeptide encoded by the portion of SEQ ID NO:X delineated by the nucleotide position nos. "Start" and "End". Also provided are polynucleotides encoding such proteins and the complementary strand thereto. The seventh and eighth columns provide the "% Identity" (percent identity) and "% Similarity" (percent similarity) observed between the aligned sequence segments of the translation product of SEQ ID NO:X and the database sequence.

The ninth column of Table 1 provides a unique "Clone ID" for a clone related to each contig sequence. This clone ID references the cDNA clone which contains at least the 5' most sequence of the assembled contig and at least a portion of SEQ ID NO:X was determined by directly sequencing the referenced clone. The reference clone may have more sequence than described in the sequence listing or the clone may have less. In the vast majority of cases, however, the clone is believed to encode a full-length polypeptide. In the case where a clone is not full-length, a full-length cDNA can be obtained by methods described elsewhere herein.

Table 3 indicates public ESTs, of which at least one, two, three, four, five, ten, or more of any one or more of these public ESTs are optionally excluded from the invention.

SEQ ID NO:X (where X may be any of the polynucleotide sequences disclosed in the sequence listing as SEQ ID NO:1 through SEQ ID NO:773) and the translated SEQ ID NO:Y (where Y may be any of the polypeptide sequences disclosed in the sequence listing as SEQ ID NO:774 through SEQ ID NO:1546) are sufficiently accurate and otherwise suitable for a

variety of uses well known in the art and described further below. For instance, SEQ ID NO:X has uses including, but not limited to, in designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the related cDNA clone contained in a library deposited with the ATCC. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling immediate applications in chromosome mapping, linkage analysis, tissue identification and/or typing, and a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y have uses that include, but are not limited to, generating antibodies which bind specifically to the colon cancer antigen polypeptides, or fragments thereof, and/or to the colon cancer antigen polypeptides encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X, the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing the related cDNA clone (deposited with the ATCC, as set forth in Table 1). The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. Further, techniques known in the art can be used to verify the nucleotide sequences of SEQ ID NO:X.

The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to vectors or plasmids which include such DNA sequences, as well as the use of the DNA sequences. The material deposited with the ATCC on:

5 **Table 2**

ATCC Deposits	Deposit Date	ATCC Designation Number
LP01, LP02, LP03, LP04, LP05, LP06, LP07, LP08, LP09, LP10, LP11,	May-20-97	209059, 209060, 209061, 209062, 209063, 209064, 209065, 209066, 209067, 209068, 209069
LP12	Jan-12-98	209579
LP13	Jan-12-98	209578
LP14	Jul-16-98	203067
LP15	Jul-16-98	203068
LP16	Feb-1-99	203609
LP17	Feb-1-99	203610
LP20	Nov-17-98	203485
LP21	Jun-18-99	PTA-252
LP22	Jun-18-99	PTA-253
LP23	Dec-22-99	PTA-1081

each is a mixture of cDNA clones derived from a variety of human tissue and cloned in either a plasmid vector or a phage vector, as shown in Table 5. These deposits are referred to as "the deposits" herein. The tissues from which the clones were derived are listed in Table 5, and the vector in which the cDNA is contained is also indicated in Table 5. The deposited material includes the cDNA clones which were partially sequenced and are related to the SEQ ID NO:X described in Table 1 (column 9). Thus, a clone which is isolatable from the ATCC Deposits by use of a sequence listed as SEQ ID NO:X may include the entire coding region of a human gene or in other cases such clone may include a substantial portion of the coding region of a human gene. Although the sequence listing lists only a portion of the DNA sequence in a clone included in the ATCC Deposits, it is well within the ability of one skilled in the art to complete the sequence of the DNA included in a clone isolatable from the

ATCC Deposits by use of a sequence (or portion thereof) listed in Table 1 by procedures hereinafter further described, and others apparent to those skilled in the art.

Also provided in Table 5 is the name of the vector which contains the cDNA clone. Each vector is routinely used in the art. The following additional information is provided for
5 convenience.

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., *Nucleic Acids Res.* 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., *Nucleic Acids Res.* 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., *Strategies* 5:58-61 (1992)) are commercially available from Stratagene
10 Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Phagemid pBS may be excised from the Lambda Zap and Uni-Zap XR vectors, and phagemid pBK may be excised from the Zap Express vector. Both phagemids may be transformed into *E. coli* strain
15 XL-1 Blue, also available from Stratagene.

Vectors pSport1, pCMVSport 1.0, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, also available from Life Technologies. See, for instance, Gruber, C. E., et al., *Focus*
20 15:59 (1993). Vector lafmid BA (Bento Soares, Columbia University, New York, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. See, for instance, Clark, J. M., *Nuc. Acids Res.*
25 16:9677-9686 (1988) and Mead, D. et al., *Bio/Technology* 9: (1991).

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, and/or the cDNA contained in a deposited cDNA clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include, but are not limited to, preparing probes or primers from the
30 disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are allelic variants, orthologs, and/or species homologs. Procedures known in the art can be used to obtain full-length genes, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologs of genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, and/or the cDNA contained in the related cDNA clone in the deposit, using information from the sequences disclosed herein or the clones deposited with the ATCC. For example, allelic variants and/or species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue.

The present invention provides a polynucleotide comprising, or alternatively consisting of, the nucleic acid sequence of SEQ ID NO:X, and/or the related cDNA clone (See, e.g., columns 1 and 9 of Table 1). The present invention also provides a polypeptide comprising, or alternatively, consisting of, the polypeptide sequence of SEQ ID NO:Y, a polypeptide encoded by SEQ ID NO:X, and/or a polypeptide encoded by the cDNA in the related cDNA clone contained in a deposited library. Polynucleotides encoding a polypeptide comprising, or alternatively consisting of, the polypeptide sequence of SEQ ID NO:Y, a polypeptide encoded by SEQ ID NO:X, and/or a polypeptide encoded by the the cDNA in the related cDNA clone contained in a deposited library, are also encompassed by the invention. The present invention further encompasses a polynucleotide comprising, or alternatively consisting of, the complement of the nucleic acid sequence of SEQ ID NO:X, and/or the complement of the coding strand of the related cDNA clone contained in a deposited library.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence would unduly burden the disclosure of this application. Accordingly, for each "Contig Id" listed in the first column of Table 3, preferably excluded are one or more polynucleotides comprising a nucleotide sequence described in the second column of Table 3 by the general formula of a-b, each of which are uniquely defined for the SEQ ID NO:X corresponding to that Contig Id in Table 1. Additionally, specific embodiments are directed to polynucleotide sequences excluding at least one, two, three, four, five, ten, or more of the specific polynucleotide sequences referenced by the Genbank Accession No. for each Contig Id which may be

included in column 3 of Table 3. In no way is this listing meant to encompass all of the sequences which may be excluded by the general formula, it is just a representative example.

Table 3.

Sequence/ Contig ID	General formula	Genbank Accession No.
500802	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 619 of SEQ ID NO:1, b is an integer of 15 to 633, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:1, and where b is greater than or equal to a + 14.	
531091	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 281 of SEQ ID NO:2, b is an integer of 15 to 295, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:2, and where b is greater than or equal to a + 14.	
553147	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 428 of SEQ ID NO:3, b is an integer of 15 to 442, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:3, and where b is greater than or equal to a + 14.	
558860	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 740 of SEQ ID NO:4, b is an integer of 15 to 754, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:4, and where b is greater than or equal to a + 14.	
561730	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 379 of SEQ ID NO:5, b is an integer of 15 to 393, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:5, and where b is greater than or equal to a + 14.	
585938	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 525 of SEQ ID NO:6, b is an integer of 15 to 539, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:6, and where b is greater than or equal to a + 14.	
587785	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 790 of SEQ ID	

	NO:214. b is an integer of 15 to 1166, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:214, and where b is greater than or equal to a + 14.	
840063	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3309 of SEQ ID NO:215, b is an integer of 15 to 3323, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:215, and where b is greater than or equal to a + 14.	
840533	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1394 of SEQ ID NO:216, b is an integer of 15 to 1408, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:216, and where b is greater than or equal to a + 14.	
840669	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2097 of SEQ ID NO:217, b is an integer of 15 to 2111, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:217, and where b is greater than or equal to a + 14.	T71029, T79145, T79226, T99989, R59589, R61735, R61734, R66190, R67070, H16201, H16200, H22960, H84137, H85574, H98850, N23572, N26340, N56614, W72249, W76334, W86530, W87654, W87653, AA057869, AA122103, AA129545, AA136524, AA137122, AA429808, AA525242, AA558970, H99223, AA584317, AA595168, AA825180, AA931521, AA938437, AI017369, N29659, N68604, W86674, AA007246
841140	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2479 of SEQ ID NO:218, b is an integer of 15 to 2493, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:218, and where b is greater than or equal to a + 14.	
841386	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1245 of SEQ ID NO:219, b is an integer of 15 to 1259, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:219, and where b is greater than or equal to a + 14.	AA429393, AA429394, AA493187, AA807096, AA836046
841480	Preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1835 of SEQ ID NO:220, b is an integer of 15 to 1849, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:220, and where b is greater than or equal to a + 14.	
841509	Preferably excluded from the present invention are	

Polynucleotide and Polypeptide Variants

The present invention is directed to variants of the polynucleotide sequence disclosed in SEQ ID NO:X or the complementary strand thereto, and/or the cDNA sequence contained in a cDNA clone contained in the deposit.

5 The present invention also encompasses variants of a colon and/or colon cancer polypeptide sequence disclosed in SEQ ID NO:Y, a polypeptide sequence encoded by the polynucleotide sequence in SEQ ID NO:X, and/or a polypeptide sequence encoded by the cDNA in the related cDNA clone contained in the deposit.

10 "Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 The present invention is also directed to nucleic acid molecules which comprise, or alternatively consist of, a nucleotide sequence which is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100%, identical to, for example, the nucleotide coding sequence in SEQ ID NO:X or the complementary strand thereto, the nucleotide coding sequence of the related cDNA contained in a deposited library or the complementary strand thereto, a nucleotide sequence encoding the polypeptide of SEQ ID NO:Y, a nucleotide sequence encoding a polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO:X, a nucleotide sequence encoding the polypeptide encoded by the cDNA in the related cDNA contained in a deposited library, and/or polynucleotide fragments of any of these nucleic acid molecules (e.g., those fragments described herein). Polypeptides encoded by these nucleic acid molecules are also encompassed by the invention. In another embodiment, the invention encompasses nucleic acid molecules which comprise or alternatively consist of, a polynucleotide which hybridizes under stringent hybridization conditions, or alternatively, under low stringency conditions, to the nucleotide coding sequence in SEQ ID NO:X, the nucleotide coding sequence of the related cDNA clone contained in a deposited library, a nucleotide sequence encoding the polypeptide of SEQ ID NO:Y, a nucleotide sequence encoding a polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO:X, a nucleotide sequence encoding the polypeptide encoded by the cDNA in the related cDNA clone contained in a deposited library, and/or polynucleotide fragments of any of these nucleic acid molecules (e.g., those fragments described herein). Polynucleotides which

20

25

30

hybridize to the complement of these nucleic acid molecules under stringent hybridization conditions or alternatively, under lower stringency conditions. are also encompassed by the invention, as are polypeptides encoded by these polynucleotides.

5 The present invention is also directed to polypeptides which comprise, or alternatively consist of, an amino acid sequence which is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to, for example, the polypeptide sequence shown in SEQ ID NO:Y, a polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO:X, a polypeptide sequence encoded by the cDNA in the related cDNA clone contained in a deposited library, and/or polypeptide fragments of any of these polypeptides (e.g., those fragments described
10 herein). Polynucleotides which hybridize to the complement of the nucleic acid molecules encoding these polypeptides under stringent hybridization conditions, or alternatively, under lower stringency conditions, are also encompassed by the invention, as are polypeptides encoded by these polynucleotides.

By a nucleic acid having a nucleotide sequence at least, for example, 95% "identical"
15 to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the nucleic acid is identical to the reference sequence except that the nucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a nucleic acid having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to
20 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be, for example, an entire sequence referred to in Table 1, an ORF (open reading frame), or any fragment specified as described herein.

25 As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global
30 sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. 6:237-245 (1990)). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be

compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other

manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that
5 the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur
10 at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence in SEQ ID
15 NO:Y or a fragment thereof, the amino acid sequence encoded by the nucleotide sequence in SEQ ID NO:X or a fragment thereof, or the amino acid sequence encoded by the cDNA in the related cDNA clone contained in a deposited library, or a fragment thereof, can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present
20 invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci.6:237- 245(1990)). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a
25 FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal
30 deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences

truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is
5 matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the
10 query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C- terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the
15 subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent
20 identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject
25 sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce
30 silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which less than 50, less

than 40, less than 30, less than 20, less than 10, or 5-50, 5-25, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II. Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level and are included in the present invention. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, as discussed herein, one or more amino acids can be deleted from the N-terminus or C-terminus of the polypeptide of the present invention without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., *J. Biotechnology* 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem.* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, as discussed herein, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more

biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show a functional activity (e.g., biological activity) of the polypeptide of the invention of which they are a variant. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity.

The present application is directed to nucleic acid molecules at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to the nucleic acid sequences disclosed herein or fragments thereof, (e.g., including but not limited to fragments encoding a polypeptide having the amino acid sequence of an N and/or C terminal deletion), irrespective of whether they encode a polypeptide having functional activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having functional activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having functional activity include, inter alia, (1) isolating a gene or allelic or splice variants thereof in a cDNA library; (2) in situ hybridization (e.g., "FISH") to metaphase chromosomal spreads to provide precise chromosomal location of the gene, as described in Verma et al., Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York (1988); and (3) Northern Blot analysis for detecting mRNA expression in specific tissues.

Preferred, however, are nucleic acid molecules having sequences at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to the nucleic acid sequences disclosed herein, which do, in fact, encode a polypeptide having a functional activity of a polypeptide of the invention.

Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to, for example, the nucleic acid sequence of the cDNA in the related cDNA clone contained in a

deposited library, the nucleic acid sequence referred to in Table 1 (SEQ ID NO:X), or fragments thereof, will encode polypeptides "having functional activity." In fact, since degenerate variants of any of these nucleotide sequences all encode the same polypeptide, in many instances, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having functional activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, *Science* 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side

chains. whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln. replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly. Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as, for example, an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a polypeptide having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course it is highly preferable for a polypeptide to have an amino acid sequence which comprises the amino acid sequence of a polypeptide of SEQ ID NO:Y, an amino acid sequence encoded by SEQ ID NO:X, and/or the amino acid sequence encoded by the cDNA in the related cDNA clone contained in a deposited library which contains, in order of ever-increasing preference, at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1

amino acid substitutions. In specific embodiments, the number of additions, substitutions, and/or deletions in the amino acid sequence of SEQ ID NO:Y or fragments thereof (e.g., the mature form and/or other fragments described herein), an amino acid sequence encoded by SEQ ID NO:X or fragments thereof, and/or the amino acid sequence encoded by the cDNA in the related cDNA clone contained in a deposited library or fragments thereof, is 1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, conservative amino acid substitutions are preferable.

Polynucleotide and Polypeptide Fragments

The present invention is also directed to polynucleotide fragments of the colon and/or colon cancer polynucleotides (nucleic acids) of the invention. In the present invention, a "polynucleotide fragment" refers, for example, to a polynucleotide having a nucleic acid sequence which: is a portion of the cDNA contained in a deposited cDNA clone; or is a portion of a polynucleotide sequence encoding the polypeptide encoded by the cDNA contained in a deposited cDNA clone; or is a portion of the polynucleotide sequence in SEQ ID NO:X or the complementary strand thereto; or is a polynucleotide sequence encoding a portion of the polypeptide of SEQ ID NO:Y; or is a polynucleotide sequence encoding a portion of a polypeptide encoded by SEQ ID NO:X or the complementary strand thereto. The nucleotide fragments of the invention are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt, at least about 50 nt, at least about 75 nt, at least about 100 nt, at least about 125 nt or at least about 150 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from, for example, the sequence contained in the cDNA in a related cDNA clone contained in a deposited library, the nucleotide sequence shown in SEQ ID NO:X or the complementary strand thereto. In this context "about" includes the particularly recited value or a value larger or smaller by several (5, 4, 3, 2, or 1) nucleotides. These nucleotide fragments have uses that include, but are not limited to, as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., at least 150, 175, 200, 250, 500, 600, 1000, or 2000 nucleotides in length) are also encompassed by the invention.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments comprising, or alternatively consisting of, a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-

400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, 2001-2050, 2051-2100, 2101-2150, 2151-2200, 2201-2250, 2251-2300, 2301-2350, 2351-2400, 2401-2450, 2451-2500, 2501-2550, 2551-2600, 2601-2650, 2651-2700, 2701-2750, 2751-2800, 2801-2850, 2851-2900, 2901-2950, 2951-3000, 3001-3050, 3051-3100, 3101-3150, 3151-3200, 3201-3250, 3251-3300, 3301-3350, 3351-3400, 3401-3450, 3451-3500, 3501-3550, 3551-3600, 3601-3650, 3651-3700, 3701-3750, 3751-3800, 3801-3850, 3851-3900, 3901-3950, 3951-4000, 4001-4050, 4051-4100, and 4101 to the end of SEQ ID NO:X, or the complementary strand thereto. In this context "about" includes the particularly recited range or a range larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has a functional activity (e.g., biological activity) of the polypeptide encoded by the polynucleotide of which the sequence is a portion. More preferably, these fragments can be used as probes or primers as discussed herein. Polynucleotides which hybridize to one or more of these nucleic acid molecules under stringent hybridization conditions or alternatively, under lower stringency conditions, are also encompassed by the invention, as are polypeptides encoded by these polynucleotides or fragments.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments comprising, or alternatively consisting of, a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, 2001-2050, 2051-2100, 2101-2150, 2151-2200, 2201-2250, 2251-2300, 2301-2350, 2351-2400, 2401-2450, 2451-2500, 2501-2550, 2551-2600, 2601-2650, 2651-2700, 2701-2750, 2751-2800, 2801-2850, 2851-2900, 2901-2950, 2951-3000, 3001-3050, 3051-3100, 3101-3150, 3151-3200, 3201-3250, 3251-3300, 3301-3350, 3351-3400, 3401-3450, 3451-3500, 3501-3550, 3551-3600, 3601-3650, 3651-3700, 3701-3750, 3751-3800, 3801-3850, 3851-3900, 3901-3950, 3951-4000, 4001-4050, 4051-4100, and 4101 to the end of the cDNA nucleotide

sequence contained in the deposited cDNA clone, or the complementary strand thereto. In this context "about" includes the particularly recited range, or a range larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has a functional activity (e.g., biological activity) of the polypeptide encoded by the cDNA nucleotide sequence contained in the deposited cDNA clone. More preferably, these fragments can be used as probes or primers as discussed herein. Polynucleotides which hybridize to one or more of these fragments under stringent hybridization conditions or alternatively, under lower stringency conditions, are also encompassed by the invention, as are polypeptides encoded by these polynucleotides or fragments.

In the present invention, a "polypeptide fragment" refers to an amino acid sequence which is a portion of that contained in SEQ ID NO:Y, a portion of an amino acid sequence encoded by the polynucleotide sequence of SEQ ID NO:X, and/or encoded by the cDNA contained in the related cDNA clone contained in a deposited library. Protein (polypeptide) fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments comprising, or alternatively consisting of, an amino acid sequence from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-320, 321-340, 341-360, 361-380, 381-400, 401-420, 421-440, 441-460, 461-480, 481-500, 501-520, 521-540, 541-560, 561-580, 581-600, 601-620, 621-640, 641-660, 661-680, 681-700, 701-720, 721-740, 741-760, 761-780, 781-800, 801-820, 821-840, 841-860, 861-880, 881-900, 901-920, 921-940, 941-960, 961-980, 981-1000, 1001-1020, 1021-1040, 1041-1060, 1061-1080, 1081-1100, 1101-1120, 1121-1140, 1141-1160, 1161-1180, 1181-1200, 1201-1220, 1221-1240, 1241-1260, 1261-1280, 1281-1300, 1301-1320, 1321-1340, 1341-1360, and 1361 to the end of SEQ ID NO:Y. Moreover, polypeptide fragments of the invention may be at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges or values, or ranges or values larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either terminus or at both termini. Polynucleotides encoding these polypeptide fragments are also encompassed by the invention.

Even if deletion of one or more amino acids from the N-terminus of a protein results in modification or loss of one or more biological functions of the protein, other functional activities (e.g., biological activities, ability to multimerize, ability to bind a ligand) may still be retained. For example, the ability of shortened muteins to induce and/or bind to antibodies which recognize the complete or mature forms of the polypeptides generally will be retained when less than the majority of the residues of the complete or mature polypeptide are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete polypeptide retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a mutein with a large number of deleted N-terminal amino acid residues may retain some biological or immunogenic activities. In fact, peptides composed of as few as six amino acid residues may often evoke an immune response.

Accordingly, polypeptide fragments of the invention include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotides encoding these polypeptide fragments are also preferred.

The present invention further provides polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence of a polypeptide disclosed herein (e.g., a polypeptide of SEQ ID NO:Y, a polypeptide encoded by the polynucleotide sequence contained in SEQ ID NO:X, and/or a polypeptide encoded by the cDNA contained in the related cDNA clone contained in a deposited library). In particular, N-terminal deletions may be described by the general formula m-q, where q is a whole integer representing the total number of amino acid residues in a polypeptide of the invention (e.g., the polypeptide disclosed in SEQ ID NO:Y), and m is defined as any integer ranging from 2 to q-6. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Also as mentioned above, even if deletion of one or more amino acids from the C-terminus of a protein results in modification or loss of one or more biological functions of

the protein, other functional activities (e.g., biological activities, ability to multimerize, ability to bind a ligand) may still be retained. For example the ability of the shortened mutein to induce and/or bind to antibodies which recognize the complete or mature forms of the polypeptide generally will be retained when less than the majority of the residues of the complete or mature polypeptide are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete polypeptide retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a mutein with a large number of deleted C-terminal amino acid residues may retain some biological or immunogenic activities. In fact, peptides composed of as few as six amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues from the carboxy terminus of the amino acid sequence of a polypeptide disclosed herein (e.g., a polypeptide of SEQ ID NO:Y, a polypeptide encoded by the polynucleotide sequence contained in SEQ ID NO:X, and/or a polypeptide encoded by the cDNA contained in the related cDNA referenced in Table 1). In particular, C-terminal deletions may be described by the general formula 1-n, where n is any whole integer ranging from 6 to q-1, and where n corresponds to the position of an amino acid residue in a polypeptide of the invention. Polynucleotides encoding these polypeptides are also encompassed by the invention.

In addition, any of the above described N- or C-terminal deletions can be combined to produce a N- and C-terminal deleted polypeptide. The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of a polypeptide encoded by SEQ ID NO:X (e.g., including, but not limited to, the preferred polypeptide disclosed as SEQ ID NO:Y), and/or the cDNA in the related cDNA clone contained in a deposited library, where n and m are integers as described above. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Any polypeptide sequence contained in the polypeptide of SEQ ID NO:Y, encoded by the polynucleotide sequences set forth as SEQ ID NO:X, or encoded by the cDNA in the related cDNA clone contained in a deposited library may be analyzed to determine certain preferred regions of the polypeptide. For example, the amino acid sequence of a polypeptide encoded by a polynucleotide sequence of SEQ ID NO:X, or the cDNA in a deposited cDNA

clone may be analyzed using the default parameters of the DNASTAR computer algorithm (DNASTAR, Inc., 1228 S. Park St., Madison, WI 53715 USA; <http://www.dnastar.com/>).

Polypeptide regions that may be routinely obtained using the DNASTAR computer algorithm include, but are not limited to, Garnier-Robson alpha-regions, beta-regions, 5 turn-regions, and coil-regions, Chou-Fasman alpha-regions, beta-regions, and turn-regions, Kyte-Doolittle hydrophilic regions and hydrophobic regions, Eisenberg alpha- and beta-amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf regions of high antigenic index. Among highly preferred polynucleotides of the invention in this regard are those that encode polypeptides comprising regions that 10 combine several structural features, such as several (e.g., 1, 2, 3 or 4) of the features set out above.

Additionally, Kyte-Doolittle hydrophilic regions and hydrophobic regions, Emini surface-forming regions, and Jameson-Wolf regions of high antigenic index (i.e., containing 15 four or more contiguous amino acids having an antigenic index of greater than or equal to 1.5, as identified using the default parameters of the Jameson-Wolf program) can routinely be used to determine polypeptide regions that exhibit a high degree of potential for antigenicity. Regions of high antigenicity are determined from data by DNASTAR analysis by choosing values which represent regions of the polypeptide which are likely to be exposed on the surface of the polypeptide in an environment in which antigen recognition may occur in the 20 process of initiation of an immune response.

Preferred polypeptide fragments of the invention are fragments comprising, or alternatively consisting of, an amino acid sequence that displays a functional activity of the polypeptide sequence of which the amino acid sequence is a fragment.

By a polypeptide demonstrating a "functional activity" is meant, a polypeptide 25 capable of displaying one or more known functional activities associated with a full-length (complete) protein of the invention. Such functional activities include, but are not limited to, biological activity, antigenicity [ability to bind (or compete with a polypeptide for binding) to an anti-polypeptide antibody], immunogenicity (ability to generate antibody which binds to a specific polypeptide of the invention), ability to form multimers with polypeptides of the 30 invention, and ability to bind to a receptor or ligand for a polypeptide.

Other preferred polypeptide fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an

activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

In preferred embodiments, polypeptides of the invention comprise, or alternatively consist of, one, two, three, four, five or more of the antigenic fragments of the polypeptide of
5 SEQ ID NO:Y, or portions thereof. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Table 4.

Sequence/ Contig ID	Predicted Epitopes
500802	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 774 as residues: Gln-1 to Ser-17, Ser-19 to Ile-25, Leu-29 to Arg-41, Ser-46 to Glu-57.
553147	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 776 as residues: Phe-1 to Ile-20.
558860	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 777 as residues: Ser-6 to Arg-11.
561730	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 778 as residues: Asn-1 to Arg-7, Leu-28 to Pro-45.
585938	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 779 as residues: Arg-10 to Ser-23, Gln-69 to His-74.
587785	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 780 as residues: Ile-1 to Ser-11, Leu-20 to Thr-30, Cys-74 to Cys-82, Leu-94 to Glu-110.
588916	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 781 as residues: Val-43 to Pro-55, Glu-92 to Ser-99.
613825	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 782 as residues: Asn-1 to Trp-11, Ser-15 to Gln-22, Ser-43 to Ala-51, Lys-58 to Gly-66.
639090	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 783 as residues: Ser-29 to Ser-35, Pro-43 to Gly-48, Gln-60 to Ser-65.
659544	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 785 as residues: Leu-10 to Glu-15, His-19 to Glu-26.
659739	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 786 as residues: Lys-70 to His-78, Lys-149 to Asn-154, Gly-209 to Leu-217, Lys-248 to Val-255, Ile-259 to Arg-264, Arg-280 to Ala-287.
661057	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 787 as residues: Cys-59 to Arg-64, Gly-110 to Asp-115, Pro-127 to Trp-132.
661313	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 788 as residues: Glu-1 to Phe-7, Lys-42 to Leu-48.
666316	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 789 as residues: Lys-27 to Asn-52.
669229	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 790 as residues: Asp-1 to Phe-12, Val-92 to Ser-103.
670471	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 791 as residues: Lys-75 to Asp-81, Glu-145 to Gln-156, Glu-163 to Arg-170, Lys-225 to Leu-231.
676611	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 792 as residues: Tyr-4 to Lys-12, Thr-23 to Asn-31, Val-52 to Thr-63, Arg-90 to Met-95.
691240	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 793 as residues: Pro-74 to Glu-79, Ser-116 to Lys-121.
702977	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 794 as residues: Pro-8 to Tyr-20.
709517	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 795 as residues: Leu-7 to Gly-12, Cys-20 to His-27.
714730	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 796 as residues: Pro-14 to Arg-23, Ala-171 to Ser-178.
714834	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 797 as residues: Ala-6 to Gly-12, Gln-18 to Arg-32.
719584	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 799 as residues: Pro-22 to Ile-31.
724637	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 800 as residues: Val-11 to Arg-34, Asn-54 to Cys-59.
728392	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 801 as

	162.
833395	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 972 as residues: Ser-3 to Gly-9.
834326	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 973 as residues: Ser-1 to Trp-19, Asn-148 to Leu-153, Tyr-235 to Trp-244.
834944	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 975 as residues: Glu-42 to Gln-51, Pro-115 to Asp-120, Arg-127 to Gly-133, Gln-199 to Gln-211.
835104	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 977 as residues: Thr-1 to Arg-14, Val-18 to Pro-23, Thr-37 to Met-44, Gln-51 to Leu-57.
835332	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 978 as residues: Thr-1 to Glu-13, Arg-135 to Asp-142, Thr-150 to Gln-155, Cys-173 to Cys-183, Cys-203 to Asp-214.
835487	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 979 as residues: Ala-13 to Arg-22, Pro-43 to Glu-57, Ala-73 to Pro-90, Arg-102 to Ser-109, Pro-114 to Gly-122, Arg-127 to Arg-138, Glu-153 to Gly-158, Pro-165 to Pro-171, Gly-185 to Arg-190, Pro-211 to Pro-216, Glu-231 to Asn-261, Ala-280 to Pro-291, Pro-303 to Gly-311, Arg-313 to Gly-326, Ala-358 to Ala-364, Pro-369 to Gly-377, Pro-390 to Gly-407, Tyr-420 to Tyr-441, Glu-461 to Thr-470, Pro-479 to Trp-487, Asp-489 to Cys-494, Gln-515 to Lys-532, Ala-572 to Asn-582, Asp-588 to Leu-594, Cys-625 to Trp-632, Tyr-639 to Arg-646.
836182	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 980 as residues: Ala-7 to Thr-17, Arg-31 to Thr-36.
836522	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 981 as residues: Gly-59 to Cys-65.
836789	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 984 as residues: Gly-18 to Gly-25, Glu-59 to Glu-64.
838577	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 985 as residues: Pro-15 to Trp-20, Pro-46 to Gln-57, Glu-68 to Phe-83.
839008	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 987 as residues: Arg-1 to Arg-13, Gln-125 to Glu-131, Asn-137 to Val-142, Gly-183 to Tyr-188, Asn-245 to Ser-251, Gln-302 to Asn-311.
840063	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 988 as residues: Gly-1 to Gly-31.
840533	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 989 as residues: Thr-16 to Pro-23, Pro-39 to Trp-48, Arg-50 to Lys-55, Gly-73 to Gly-79.
840669	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 990 as residues: Met-27 to Gln-33, Gln-49 to Gly-56, Thr-63 to Leu-70, Thr-115 to Arg-127, Pro-174 to Asn-184.
841140	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 991 as residues: Arg-17 to Phe-24, Pro-113 to Gly-121, Thr-235 to Met-240.
841386	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 992 as residues: Val-58 to Met-66, Pro-134 to Lys-143, Tyr-163 to Ala-170, Val-178 to Lys-187, Pro-207 to Gly-212.
841900	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 996 as residues: Ile-2 to Phe-12.
842054	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 997 as residues: Asp-27 to Trp-32, Pro-89 to Glu-99, Arg-112 to Lys-123.
843061	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 998 as residues: Leu-3 to Gly-18, His-36 to His-57, Lys-136 to Leu-145, Gly-174 to Trp-184, Lys-188 to Tyr-196, Lys-204 to Asp-211, Pro-293 to Ser-305, Glu-321 to Asp-333, Gly-342 to Lys-348, Ala-371 to Asp-377, Asp-439 to Leu-449, Ala-521 to Gly-529, Tyr-583 to Trp-599, Asn-639 to Ser-644, Leu-738 to Leu-745.
843544	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 999 as residues: Tyr-11 to Phe-18, Ser-34 to Lys-43.
844092	Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 1000 as

The present invention encompasses polypeptides comprising, or alternatively consisting of, an epitope of the polypeptide sequence shown in SEQ ID NO:Y, or an epitope of the polypeptide sequence encoded by the cDNA in the related cDNA clone contained in a deposited library or encoded by a polynucleotide that hybridizes to the complement of an epitope encoding sequence of SEQ ID NO:X, or an epitope encoding sequence contained in the deposited cDNA clone under stringent hybridization conditions, or alternatively, under lower stringency hybridization conditions, as defined supra. The present invention further encompasses polynucleotide sequences encoding an epitope of a polypeptide sequence of the invention (such as, for example, the sequence disclosed in SEQ ID NO:X), polynucleotide sequences of the complementary strand of a polynucleotide sequence encoding an epitope of the invention, and polynucleotide sequences which hybridize to this complementary strand under stringent hybridization conditions or alternatively, under lower stringency hybridization conditions, as defined supra.

The term "epitopes," as used herein, refers to portions of a polypeptide having antigenic or immunogenic activity in an animal, preferably a mammal, and most preferably in a human. In a preferred embodiment, the present invention encompasses a polypeptide comprising an epitope, as well as the polynucleotide encoding this polypeptide. An "immunogenic epitope," as used herein, is defined as a portion of a protein that elicits an antibody response in an animal, as determined by any method known in the art, for example, by the methods for generating antibodies described infra. (See, for example, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983)). The term "antigenic epitope," as used herein, is defined as a portion of a protein to which an antibody can immunospecifically bind its antigen as determined by any method well known in the art, for example, by the immunoassays described herein. Immunospecific binding excludes non-specific binding but does not necessarily exclude cross-reactivity with other antigens. Antigenic epitopes need not necessarily be immunogenic.

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least 4, at least 5, at least 6, at least 7, more preferably at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 20, at least 25, at least 30, at least 40, at

least 50, and, most preferably, between about 15 to about 30 amino acids. Preferred polypeptides comprising immunogenic or antigenic epitopes are at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acid residues in length. Additional non-exclusive preferred antigenic epitopes include the antigenic epitopes disclosed herein, as well as portions thereof. Antigenic epitopes are useful, for example, to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. Preferred antigenic epitopes include the antigenic epitopes disclosed herein, as well as any combination of two, three, four, five or more of these antigenic epitopes. Antigenic epitopes can be used as the target molecules in immunoassays. (See, for instance, Wilson et al., Cell 10 37:767-778 (1984); Sutcliffe et al., Science 219:660-666 (1983)).

Similarly, immunogenic epitopes can be used, for example, to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle et al., J. Gen. Virol. 66:2347-2354 (1985)). Preferred immunogenic epitopes include the immunogenic epitopes disclosed herein, as well as any combination of two, three, four, five or more of these immunogenic epitopes. The polypeptides comprising one or more immunogenic epitopes may be presented for eliciting an antibody response together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse), or, if the polypeptide is of sufficient length (at least about 25 amino acids), the polypeptide may be presented without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting).

Epitope-bearing polypeptides of the present invention may be used to induce antibodies according to methods well known in the art including, but not limited to, in vivo immunization, in vitro immunization, and phage display methods. See, e.g., Sutcliffe et al., supra; Wilson et al., supra, and Bittle et al., J. Gen. Virol., 66:2347-2354 (1985). If in vivo immunization is used, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by coupling the peptide to a macromolecular carrier, such as keyhole limpet hemacyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine residues may be coupled to a carrier using a linker such as maleimidobenzoyl- N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carriers using a more general linking agent such as glutaraldehyde. Animals such as rabbits, rats and mice

are immunized with either free or carrier- coupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 µg of peptide or carrier protein and Freund's adjuvant or any other adjuvant known for stimulating an immune response. Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide antibody which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid support and elution of the selected antibodies according to methods well known in the art.

As one of skill in the art will appreciate, and as discussed above, the polypeptides of the present invention, and immunogenic and/or antigenic epitope fragments thereof can be fused to other polypeptide sequences. For example, the polypeptides of the present invention may be fused with the constant domain of immunoglobulins (IgA, IgE, IgG, IgM), or portions thereof (CH1, CH2, CH3, or any combination thereof and portions thereof) resulting in chimeric polypeptides. Such fusion proteins may facilitate purification and may increase half-life in vivo. This has been shown for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. See, e.g., EP 394,827; Traunecker et al., *Nature*, 331:84-86 (1988). Enhanced delivery of an antigen across the epithelial barrier to the immune system has been demonstrated for antigens (e.g., insulin) conjugated to an FcRn binding partner such as IgG or Fc fragments (see, e.g., PCT Publications WO 96/22024 and WO 99/04813). IgG Fusion proteins that have a disulfide-linked dimeric structure due to the IgG portion disulfide bonds have also been found to be more efficient in binding and neutralizing other molecules than monomeric polypeptides or fragments thereof alone. See, e.g., Fountoulakis et al., *J. Biochem.*, 270:3958-3964 (1995).

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, may be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for

immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol. Chem.* 270:9459-9471 (1995).)

5 Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et
10 al., *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., *Cell* 37:767 (1984).)

15 Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

 Nucleic acids encoding the above epitopes can also be recombined with a gene of interest as an epitope tag (e.g., the hemagglutinin ("HA") tag or flag tag) to aid in detection and purification of the expressed polypeptide. For example, a system described by Janknecht et al. allows for the ready purification of non-denatured fusion proteins expressed
20 in human cell lines (Janknecht et al., *Proc. Natl. Acad. Sci. USA* 88:8972- 897 (1991)). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the open reading frame of the gene is translationally fused to an amino-terminal tag consisting of six histidine residues. The tag serves as a matrix binding domain for the fusion protein. Extracts from cells infected with the recombinant vaccinia virus are loaded onto
25 Ni²⁺ nitriloacetic acid-agarose column and histidine-tagged proteins can be selectively eluted with imidazole-containing buffers.

 Additional fusion proteins of the invention may be generated through the techniques of gene-shuffling, motif-shuffling, exon-shuffling, and/or codon-shuffling (collectively referred to as "DNA shuffling"). DNA shuffling may be employed to modulate the activities
30 of polypeptides of the invention, such methods can be used to generate polypeptides with altered activity, as well as agonists and antagonists of the polypeptides. See, generally, U.S. Patent Nos. 5,605,793; 5,811,238; 5,830,721; 5,834,252; and 5,837,458, and Patten et al.,

Curr. Opin. Biotechnol. 8:724-33 (1997); Harayama, Trends Biotechnol. 16(2):76-82 (1998); Hansson, et al., J. Mol. Biol. 287:265-76 (1999); and Lorenzo and Blasco, Biotechniques 24(2):308-13 (1998) (each of these patents and publications are hereby incorporated by reference in its entirety). In one embodiment, alteration of polynucleotides corresponding to SEQ ID NO:X and the polypeptides encoded by these polynucleotides may be achieved by DNA shuffling. DNA shuffling involves the assembly of two or more DNA segments by homologous or site-specific recombination to generate variation in the polynucleotide sequence. In another embodiment, polynucleotides of the invention, or the encoded polypeptides, may be altered by being subjected to random mutagenesis by error-prone PCR, random nucleotide insertion or other methods prior to recombination. In another embodiment, one or more components, motifs, sections, parts, domains, fragments, etc., of a polynucleotide encoding a polypeptide of the invention may be recombined with one or more components, motifs, sections, parts, domains, fragments, etc. of one or more heterologous molecules.

As discussed herein, any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, polypeptides of the present invention which are shown to be secreted can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

In certain preferred embodiments, proteins of the invention comprise fusion proteins wherein the polypeptides are N and/or C-terminal deletion mutants. In preferred embodiments, the application is directed to nucleic acid molecules at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequences encoding polypeptides having the amino acid sequence of the specific N- and C-terminal deletions mutants. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

10 Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides of the invention may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include,

but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* or *Pichia pastoris* (ATCC Accession No. 201178)); insect cells such as *Drosophila* S2 and *Spodoptera Sf9* cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells.

5 Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-
10 3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Preferred expression vectors for use in yeast systems include, but are not limited to pYES2, pYD1, pTEF1/Zeo, pYES2/GS, pPICZ, pGAPZ, pGAPZalph, pPIC9, pPIC3.5, pHIL-D2, pHIL-S1,
15 pPIC3.5K, pPIC9K, and PAO815 (all available from Invitrogen, Carlsbad, CA). Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in
20 many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid
25 extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention can also be recovered from: products purified
30 from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast,

higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

In one embodiment, the yeast *Pichia pastoris* is used to express polypeptides of the invention in a eukaryotic system. *Pichia pastoris* is a methylotrophic yeast which can metabolize methanol as its sole carbon source. A main step in the methanol metabolism pathway is the oxidation of methanol to formaldehyde using O₂. This reaction is catalyzed by the enzyme alcohol oxidase. In order to metabolize methanol as its sole carbon source, *Pichia pastoris* must generate high levels of alcohol oxidase due, in part, to the relatively low affinity of alcohol oxidase for O₂. Consequently, in a growth medium depending on methanol as a main carbon source, the promoter region of one of the two alcohol oxidase genes (*AOX1*) is highly active. In the presence of methanol, alcohol oxidase produced from the *AOX1* gene comprises up to approximately 30% of the total soluble protein in *Pichia pastoris*. See, Ellis, S.B., *et al.*, *Mol. Cell. Biol.* 5:1111-21 (1985); Koutz, P.J., *et al.*, *Yeast* 5:167-77 (1989); Tschopp, J.F., *et al.*, *Nucl. Acids Res.* 15:3859-76 (1987). Thus, a heterologous coding sequence, such as, for example, a polynucleotide of the present invention, under the transcriptional regulation of all or part of the *AOX1* regulatory sequence is expressed at exceptionally high levels in *Pichia* yeast grown in the presence of methanol.

In one example, the plasmid vector pPIC9K is used to express DNA encoding a polypeptide of the invention, as set forth herein, in a *Pichea* yeast system essentially as described in "*Pichia* Protocols: Methods in Molecular Biology," D.R. Higgins and J. Cregg, eds. The Humana Press, Totowa, NJ, 1998. This expression vector allows expression and secretion of a polypeptide of the invention by virtue of the strong *AOX1* promoter linked to

the *Pichia pastoris* alkaline phosphatase (PHO) secretory signal peptide (i.e., leader) located upstream of a multiple cloning site.

Many other yeast vectors could be used in place of pPIC9K, such as, pYES2, pYD1, pTEF1/Zeo, pYES2/GS, pPICZ, pGAPZ, pGAPZalpha, pPIC9, pPIC3.5, pHIL-D2, pHIL-S1, pPIC3.5K, and PAO815. as one skilled in the art would readily appreciate, as long as the proposed expression construct provides appropriately located signals for transcription, translation, secretion (if desired), and the like, including an in-frame AUG as required.

In another embodiment, high-level expression of a heterologous coding sequence, such as, for example, a polynucleotide of the present invention, may be achieved by cloning the heterologous polynucleotide of the invention into an expression vector such as, for example, pGAPZ or pGAPZalpha, and growing the yeast culture in the absence of methanol.

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly mammalian origin, that have been engineered to delete or replace endogenous genetic material (e.g., coding sequence), and/or to include genetic material (e.g., heterologous polynucleotide sequences) that is operably associated with polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g., promoter and/or enhancer) and endogenous polynucleotide sequences via homologous recombination (see, e.g., U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller et al., Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989); and Zijlstra et al., Nature 342:435-438 (1989), the disclosures of each of which are incorporated by reference in their entireties).

In addition, polypeptides of the invention can be chemically synthesized using techniques known in the art (e.g., see Creighton, 1983, Proteins: Structures and Molecular Principles, W.H. Freeman & Co., N.Y., and Hunkapiller et al., Nature, 310:105-111 (1984)). For example, a polypeptide corresponding to a fragment of a polypeptide can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the

polypeptide sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ -Abu, ϵ -Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β -alanine, fluoro-amino acids, designer amino acids such as β -methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

- 10 Non-naturally occurring variants may be produced using art-known mutagenesis techniques, which include, but are not limited to oligonucleotide mediated mutagenesis, alanine scanning, PCR mutagenesis, site directed mutagenesis (see, e.g., Carter *et al.*, *Nucl. Acids Res.* 13:4331 (1986); and Zoller *et al.*, *Nucl. Acids Res.* 10:6487 (1982)), cassette mutagenesis (see, e.g., Wells *et al.*, *Gene* 34:315 (1985)), restriction selection mutagenesis (see, e.g., Wells *et al.*, *Philos. Trans. R. Soc. London SerA* 317:415 (1986)).

- 15 The invention additionally, encompasses polypeptides of the present invention which are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to 20 specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH_4 , acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; etc.

- 25 Additional post-translational modifications encompassed by the invention include, for example, e.g., N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of procaryotic host cell expression. The polypeptides may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the protein.

30 Also provided by the invention are chemically modified derivatives of the polypeptides of the invention which may provide additional advantages such as increased

solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (see U.S. Patent No. 4,179,337). The chemical moieties for derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (e.g., the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog). For example, the polyethylene glycol may have an average molecular weight of about 200; 500; 1000; 1500; 2000; 2500; 3000; 3500; 4000; 4500; 5000; 5500; 6000; 6500; 7000; 7500; 8000; 8500; 9000; 9500; 10,000; 10,500; 11,000; 11,500; 12,000; 12,500; 13,000; 13,500; 14,000; 14,500; 15,000; 15,500; 16,000; 16,500; 17,000; 17,500; 18,000; 18,500; 19,000; 19,500; 20,000; 25,000; 30,000; 35,000; 40,000; 50,000; 55,000; 60,000; 65,000; 70,000; 75,000; 80,000; 85,000; 90,000; 95,000; or 100,000 kDa.

As noted above, the polyethylene glycol may have a branched structure. Branched polyethylene glycols are described, for example, in U.S. Patent No. 5,643,575; Morpurgo *et al.*, *Appl. Biochem. Biotechnol.* 56:59-72 (1996); Vorobjev *et al.*, *Nucleosides Nucleotides* 18:2745-2750 (1999); and Caliceti *et al.*, *Bioconjug. Chem.* 10:638-646 (1999), the disclosures of each of which are incorporated herein by reference.

The polyethylene glycol molecules (or other chemical moieties) should be attached to the protein with consideration of effects on functional or antigenic domains of the protein. There are a number of attachment methods available to those skilled in the art, e.g., EP 0 401 384, herein incorporated by reference (coupling PEG to G-CSF), see also Malik *et al.*, *Exp. Hematol.* 20:1028-1035 (1992) (reporting pegylation of GM-CSF using tresyl chloride). For example, polyethylene glycol may be covalently bound through amino acid residues via a

reactive group, such as, a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues: those having a free carboxyl group may include aspartic acid residues glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. Preferred for therapeutic purposes is attachment at an amino group, such as attachment at the N-terminus or lysine group.

As suggested above, polyethylene glycol may be attached to proteins via linkage to any of a number of amino acid residues. For example, polyethylene glycol can be linked to proteins via covalent bonds to lysine, histidine, aspartic acid, glutamic acid, or cysteine residues. One or more reaction chemistries may be employed to attach polyethylene glycol to specific amino acid residues (e.g., lysine, histidine, aspartic acid, glutamic acid, or cysteine) of the protein or to more than one type of amino acid residue (e.g., lysine, histidine, aspartic acid, glutamic acid, cysteine and combinations thereof) of the protein.

One may specifically desire proteins chemically modified at the N-terminus. Using polyethylene glycol as an illustration of the present composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, etc.), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus modification may be accomplished by reductive alkylation which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

As indicated above, pegylation of the proteins of the invention may be accomplished by any number of means. For example, polyethylene glycol may be attached to the protein either directly or by an intervening linker. Linkerless systems for attaching polyethylene glycol to proteins are described in Delgado *et al.*, *Crit. Rev. Thera. Drug Carrier Sys.* 9:249-

304 (1992); Francis *et al.*, *Intern. J. of Hematol.* 68:1-18 (1998); U.S. Patent No. 4,002,531; U.S. Patent No. 5,349,052; WO 95/06058; and WO 98/32466, the disclosures of each of which are incorporated herein by reference.

One system for attaching polyethylene glycol directly to amino acid residues of proteins without an intervening linker employs tresylated MPEG, which is produced by the modification of monmethoxy polyethylene glycol (MPEG) using tresylchloride ($\text{ClSO}_2\text{CH}_2\text{CF}_3$). Upon reaction of protein with tresylated MPEG, polyethylene glycol is directly attached to amine groups of the protein. Thus, the invention includes protein-polyethylene glycol conjugates produced by reacting proteins of the invention with a polyethylene glycol molecule having a 2,2,2-trifluoroethane sulphonyl group.

Polyethylene glycol can also be attached to proteins using a number of different intervening linkers. For example, U.S. Patent No. 5,612,460, the entire disclosure of which is incorporated herein by reference, discloses urethane linkers for connecting polyethylene glycol to proteins. Protein-polyethylene glycol conjugates wherein the polyethylene glycol is attached to the protein by a linker can also be produced by reaction of proteins with compounds such as MPEG-succinimidylsuccinate, MPEG activated with 1,1'-carbonyldiimidazole, MPEG-2,4,5-trichloropenylcarbonate, MPEG-p-nitrophenolcarbonate, and various MPEG-succinate derivatives. A number additional polyethylene glycol derivatives and reaction chemistries for attaching polyethylene glycol to proteins are described in WO 98/32466, the entire disclosure of which is incorporated herein by reference. Pegylated protein products produced using the reaction chemistries set out herein are included within the scope of the invention.

The number of polyethylene glycol moieties attached to each protein of the invention (*i.e.*, the degree of substitution) may also vary. For example, the pegylated proteins of the invention may be linked, on average, to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 17, 20, or more polyethylene glycol molecules. Similarly, the average degree of substitution within ranges such as 1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11, 10-12, 11-13, 12-14, 13-15, 14-16, 15-17, 16-18, 17-19, or 18-20 polyethylene glycol moieties per protein molecule. Methods for determining the degree of substitution are discussed, for example, in Delgado *et al.*, *Crit. Rev. Thera. Drug Carrier Sys.* 9:249-304 (1992).

The colon cancer antigen polypeptides of the invention may be in monomers or multimers (*i.e.*, dimers, trimers, tetramers and higher multimers). Accordingly, the present

invention relates to monomers and multimers of the polypeptides of the invention, their preparation, and compositions (preferably, Therapeutics) containing them. In specific embodiments, the polypeptides of the invention are monomers, dimers, trimers or tetramers. In additional embodiments, the multimers of the invention are at least dimers, at least trimers, or at least tetramers.

Multimers encompassed by the invention may be homomers or heteromers. As used herein, the term homomer, refers to a multimer containing only polypeptides corresponding to the amino acid sequence of SEQ ID NO:Y or an amino acid sequence encoded by SEQ ID NO:X, and/or an amino acid sequence encoded by the cDNA in a related cDNA clone contained in a deposited library (including fragments, variants, splice variants, and fusion proteins, corresponding to any one of these as described herein). These homomers may contain polypeptides having identical or different amino acid sequences. In a specific embodiment, a homomer of the invention is a multimer containing only polypeptides having an identical amino acid sequence. In another specific embodiment, a homomer of the invention is a multimer containing polypeptides having different amino acid sequences. In specific embodiments, the multimer of the invention is a homodimer (e.g., containing polypeptides having identical or different amino acid sequences) or a homotrimer (e.g., containing polypeptides having identical and/or different amino acid sequences). In additional embodiments, the homomeric multimer of the invention is at least a homodimer, at least a homotrimer, or at least a homotetramer.

As used herein, the term heteromer refers to a multimer containing one or more heterologous polypeptides (i.e., polypeptides of different proteins) in addition to the polypeptides of the invention. In a specific embodiment, the multimer of the invention is a heterodimer, a heterotrimer, or a heterotetramer. In additional embodiments, the heteromeric multimer of the invention is at least a heterodimer, at least a heterotrimer, or at least a heterotetramer.

Multimers of the invention may be the result of hydrophobic, hydrophilic, ionic and/or covalent associations and/or may be indirectly linked, by for example, liposome formation. Thus, in one embodiment, multimers of the invention, such as, for example, homodimers or homotrimers, are formed when polypeptides of the invention contact one another in solution. In another embodiment, heteromultimers of the invention, such as, for example, heterotrimers or heterotetramers, are formed when polypeptides of the invention

contact antibodies to the polypeptides of the invention (including antibodies to the heterologous polypeptide sequence in a fusion protein of the invention) in solution. In other embodiments, multimers of the invention are formed by covalent associations with and/or between the polypeptides of the invention. Such covalent associations may involve one or more amino acid residues contained in the polypeptide sequence (e.g., that recited in SEQ ID NO:Y, or contained in a polypeptide encoded by SEQ ID NO:X, and/or by the cDNA in the related cDNA clone contained in a deposited library). In one instance, the covalent associations are cross-linking between cysteine residues located within the polypeptide sequences which interact in the native (i.e., naturally occurring) polypeptide. In another instance, the covalent associations are the consequence of chemical or recombinant manipulation. Alternatively, such covalent associations may involve one or more amino acid residues contained in the heterologous polypeptide sequence in a fusion protein. In one example, covalent associations are between the heterologous sequence contained in a fusion protein of the invention (see, e.g., US Patent Number 5,478,925). In a specific example, the covalent associations are between the heterologous sequence contained in a Fc fusion protein of the invention (as described herein). In another specific example, covalent associations of fusion proteins of the invention are between heterologous polypeptide sequence from another protein that is capable of forming covalently associated multimers, such as for example, osteoprotegerin (see, e.g., International Publication NO: WO 98/49305, the contents of which are herein incorporated by reference in its entirety). In another embodiment, two or more polypeptides of the invention are joined through peptide linkers. Examples include those peptide linkers described in U.S. Pat. No. 5,073,627 (hereby incorporated by reference). Proteins comprising multiple polypeptides of the invention separated by peptide linkers may be produced using conventional recombinant DNA technology.

Another method for preparing multimer polypeptides of the invention involves use of polypeptides of the invention fused to a leucine zipper or isoleucine zipper polypeptide sequence. Leucine zipper and isoleucine zipper domains are polypeptides that promote multimerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins (Landschulz et al., Science 240:1759, (1988)), and have since been found in a variety of different proteins. Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize. Examples of leucine zipper domains suitable for producing soluble multimeric proteins of the

invention are those described in PCT application WO 94/10308, hereby incorporated by reference. Recombinant fusion proteins comprising a polypeptide of the invention fused to a polypeptide sequence that dimerizes or trimerizes in solution are expressed in suitable host cells, and the resulting soluble multimeric fusion protein is recovered from the culture supernatant using techniques known in the art.

Trimeric polypeptides of the invention may offer the advantage of enhanced biological activity. Preferred leucine zipper moieties and isoleucine moieties are those that preferentially form trimers. One example is a leucine zipper derived from lung surfactant protein D (SPD), as described in Hoppe et al. (FEBS Letters 344:191, (1994)) and in U.S. patent application Ser. No. 08/446,922, hereby incorporated by reference. Other peptides derived from naturally occurring trimeric proteins may be employed in preparing trimeric polypeptides of the invention.

In another example, proteins of the invention are associated by interactions between Flag® polypeptide sequence contained in fusion proteins of the invention containing Flag® polypeptide sequence. In a further embodiment, associations proteins of the invention are associated by interactions between heterologous polypeptide sequence contained in Flag® fusion proteins of the invention and anti-Flag® antibody.

The multimers of the invention may be generated using chemical techniques known in the art. For example, polypeptides desired to be contained in the multimers of the invention may be chemically cross-linked using linker molecules and linker molecule length optimization techniques known in the art (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Additionally, multimers of the invention may be generated using techniques known in the art to form one or more inter-molecule cross-links between the cysteine residues located within the sequence of the polypeptides desired to be contained in the multimer (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Further, polypeptides of the invention may be routinely modified by the addition of cysteine or biotin to the C-terminus or N-terminus of the polypeptide and techniques known in the art may be applied to generate multimers containing one or more of these modified polypeptides (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Additionally, techniques known in the art may be applied to generate liposomes containing the polypeptide

components desired to be contained in the multimer of the invention (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety).

Alternatively, multimers of the invention may be generated using genetic engineering techniques known in the art. In one embodiment, polypeptides contained in multimers of the invention are produced recombinantly using fusion protein technology described herein or otherwise known in the art (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). In a specific embodiment, polynucleotides coding for a homodimer of the invention are generated by ligating a polynucleotide sequence encoding a polypeptide of the invention to a sequence encoding a linker polypeptide and then further to a synthetic polynucleotide encoding the translated product of the polypeptide in the reverse orientation from the original C-terminus to the N-terminus (lacking the leader sequence) (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). In another embodiment, recombinant techniques described herein or otherwise known in the art are applied to generate recombinant polypeptides of the invention which contain a transmembrane domain (or hydrophobic or signal peptide) and which can be incorporated by membrane reconstitution techniques into liposomes (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety).

Antibodies

Further polypeptides of the invention relate to antibodies and T-cell antigen receptors (TCR) which immunospecifically bind a polypeptide, polypeptide fragment, or variant of SEQ ID NO:Y, and/or an epitope, of the present invention (as determined by immunoassays well known in the art for assaying specific antibody-antigen binding). Antibodies of the invention include, but are not limited to, polyclonal, monoclonal, multispecific, human, humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab') fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies (including, e.g., anti-Id antibodies to antibodies of the invention), and epitope-binding fragments of any of the above. The term "antibody," as used herein, refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that immunospecifically binds an antigen. The immunoglobulin molecules of the invention can be of any type (e.g., IgG,

IgE, IgM, IgD, IgA and IgY), class (e.g., IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2) or subclass of immunoglobulin molecule.

Most preferably the antibodies are human antigen-binding antibody fragments of the present invention and include, but are not limited to, Fab, Fab' and F(ab')₂, Fd, single-chain Fvs (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv) and fragments comprising either a VL or VH domain. Antigen-binding antibody fragments, including single-chain antibodies, may comprise the variable region(s) alone or in combination with the entirety or a portion of the following: hinge region, CH1, CH2, and CH3 domains. Also included in the invention are antigen-binding fragments also comprising any combination of variable region(s) with a hinge region, CH1, CH2, and CH3 domains. The antibodies of the invention may be from any animal origin including birds and mammals. Preferably, the antibodies are human, murine (e.g., mouse and rat), donkey, sheep, rabbit, goat, guinea pig, camel, horse, or chicken. As used herein, "human" antibodies include antibodies having the amino acid sequence of a human immunoglobulin and include antibodies isolated from human immunoglobulin libraries or from animals transgenic for one or more human immunoglobulin and that do not express endogenous immunoglobulins, as described infra and, for example in, U.S. Patent No. 5,939,598 by Kucherlapati et al.

The antibodies of the present invention may be monospecific, bispecific, trispecific or of greater multispecificity. Multispecific antibodies may be specific for different epitopes of a polypeptide of the present invention or may be specific for both a polypeptide of the present invention as well as for a heterologous epitope, such as a heterologous polypeptide or solid support material. See, e.g., PCT publications WO 93/17715; WO 92/08802; WO 91/00360; WO 92/05793; Tutt, et al., J. Immunol. 147:60-69 (1991); U.S. Patent Nos. 4,474,893; 4,714,681; 4,925,648; 5,573,920; 5,601,819; Kostelny et al., J. Immunol. 148:1547-1553 (1992).

Antibodies of the present invention may be described or specified in terms of the epitope(s) or portion(s) of a polypeptide of the present invention which they recognize or specifically bind. The epitope(s) or polypeptide portion(s) may be specified as described herein, e.g., by N-terminal and C-terminal positions, or by size in contiguous amino acid residues. Antibodies which specifically bind any epitope or polypeptide of the present invention may also be excluded. Therefore, the present invention includes antibodies that

specifically bind polypeptides of the present invention, and allows for the exclusion of the same.

Antibodies of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies that do not bind any other analog, ortholog, or homolog of a polypeptide of the present invention are included. Antibodies that bind polypeptides with at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, and at least 50% identity (as calculated using methods known in the art and described herein) to a polypeptide of the present invention are also included in the present invention. In specific embodiments, antibodies of the present invention cross-react with murine, rat and/or rabbit homologs of human proteins and the corresponding epitopes thereof. Antibodies that do not bind polypeptides with less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, and less than 50% identity (as calculated using methods known in the art and described herein) to a polypeptide of the present invention are also included in the present invention. In a specific embodiment, the above-described cross-reactivity is with respect to any single specific antigenic or immunogenic polypeptide, or combination(s) of 2, 3, 4, 5, or more of the specific antigenic and/or immunogenic polypeptides disclosed herein. Further included in the present invention are antibodies which bind polypeptides encoded by polynucleotides which hybridize to a polynucleotide of the present invention under stringent hybridization conditions (as described herein). Antibodies of the present invention may also be described or specified in terms of their binding affinity to a polypeptide of the invention. Preferred binding affinities include those with a dissociation constant or K_d less than 5×10^{-2} M, 10^{-2} M, 5×10^{-3} M, 10^{-3} M, 5×10^{-4} M, 10^{-4} M, 5×10^{-5} M, 10^{-5} M, 5×10^{-6} M, 10^{-6} M, 5×10^{-7} M, 10^{-7} M, 5×10^{-8} M, 10^{-8} M, 5×10^{-9} M, 10^{-9} M, 5×10^{-10} M, 10^{-10} M, 5×10^{-11} M, 10^{-11} M, 5×10^{-12} M, 10^{-12} M, 5×10^{-13} M, 10^{-13} M, 5×10^{-14} M, 10^{-14} M, 5×10^{-15} M, or 10^{-15} M.

The invention also provides antibodies that competitively inhibit binding of an antibody to an epitope of the invention as determined by any method known in the art for determining competitive binding, for example, the immunoassays described herein. In preferred embodiments, the antibody competitively inhibits binding to the epitope by at least 95%, at least 90%, at least 85 %, at least 80%, at least 75%, at least 70%, at least 60%, or at least 50%.

Antibodies of the present invention may act as agonists or antagonists of the polypeptides of the present invention. For example, the present invention includes antibodies which disrupt the receptor/ligand interactions with the polypeptides of the invention either partially or fully. Preferably, antibodies of the present invention bind an antigenic epitope disclosed herein, or a portion thereof. The invention features both receptor-specific antibodies and ligand-specific antibodies. The invention also features receptor-specific antibodies which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. For example, receptor activation can be determined by detecting the phosphorylation (e.g., tyrosine or serine/threonine) of the receptor or its substrate by immunoprecipitation followed by western blot analysis (for example, as described supra). In specific embodiments, antibodies are provided that inhibit ligand activity or receptor activity by at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 60%, or at least 50% of the activity in absence of the antibody.

The invention also features receptor-specific antibodies which both prevent ligand binding and receptor activation as well as antibodies that recognize the receptor-ligand complex, and, preferably, do not specifically recognize the unbound receptor or the unbound ligand. Likewise, included in the invention are neutralizing antibodies which bind the ligand and prevent binding of the ligand to the receptor, as well as antibodies which bind the ligand, thereby preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included in the invention are antibodies which activate the receptor. These antibodies may act as receptor agonists, i.e., potentiate or activate either all or a subset of the biological activities of the ligand-mediated receptor activation, for example, by inducing dimerization of the receptor. The antibodies may be specified as agonists, antagonists or inverse agonists for biological activities comprising the specific biological activities of the peptides of the invention disclosed herein. The above antibody agonists can be made using methods known in the art. See, e.g., PCT publication WO 96/40281; U.S. Patent No. 5,811,097; Deng et al., *Blood* 92(6):1981-1988 (1998); Chen et al., *Cancer Res.* 58(16):3668-3678 (1998); Harrop et al., *J. Immunol.* 161(4):1786-1794 (1998); Zhu et al., *Cancer Res.* 58(15):3209-3214 (1998); Yoon et al., *J. Immunol.* 160(7):3170-3179 (1998); Prat et al., *J. Cell. Sci.* 111(Pt2):237-247 (1998); Pitard et al., *J. Immunol. Methods* 205(2):177-190 (1997); Liautard et al., *Cytokine* 9(4):233-241 (1997); Carlson et al., *J. Biol.*

Chem. 272(17):11295-11301 (1997); Taryman et al., Neuron 14(4):755-762 (1995); Muller et al., Structure 6(9):1153-1167 (1998); Bartunek et al., Cytokine 8(1):14-20 (1996) (which are all incorporated by reference herein in their entireties).

Antibodies of the present invention may be used, for example, but not limited to, to
5 purify, detect, and target the polypeptides of the present invention, including both in vitro and
in vivo diagnostic and therapeutic methods. For example, the antibodies have use in
immunoassays for qualitatively and quantitatively measuring levels of the polypeptides of the
present invention in biological samples. See, e.g., Harlow et al., Antibodies: A Laboratory
Manual. (Cold Spring Harbor Laboratory Press, 2nd ed. 1988) (incorporated by reference
10 herein in its entirety).

As discussed in more detail below, the antibodies of the present invention may be
used either alone or in combination with other compositions. The antibodies may further be
recombinantly fused to a heterologous polypeptide at the N- or C-terminus or chemically
conjugated (including covalently and non-covalently conjugations) to polypeptides or other
15 compositions. For example, antibodies of the present invention may be recombinantly fused
or conjugated to molecules useful as labels in detection assays and effector molecules such as
heterologous polypeptides, drugs, radionuclides, or toxins. See, e.g., PCT publications WO
92/08495; WO 91/14438; WO 89/12624; U.S. Patent No. 5,314,995; and EP 396,387.

The antibodies of the invention include derivatives that are modified, i.e., by the
20 covalent attachment of any type of molecule to the antibody such that covalent attachment
does not prevent the antibody from generating an anti-idiotypic response. For example, but
not by way of limitation, the antibody derivatives include antibodies that have been modified,
e.g., by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by
known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other
25 protein, etc. Any of numerous chemical modifications may be carried out by known
techniques, including, but not limited to specific chemical cleavage, acetylation, formylation,
metabolic synthesis of tunicamycin, etc. Additionally, the derivative may contain one or
more non-classical amino acids.

The antibodies of the present invention may be generated by any suitable method
30 known in the art. Polyclonal antibodies to an antigen-of-interest can be produced by various
procedures well known in the art. For example, a polypeptide of the invention can be
administered to various host animals including, but not limited to, rabbits, mice, rats, etc. to

induce the production of sera containing polyclonal antibodies specific for the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species, and include but are not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum. Such adjuvants are also well known in the art.

Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technologies, or a combination thereof. For example, monoclonal antibodies can be produced using hybridoma techniques including those known in the art and taught, for example, in Harlow et al., *Antibodies: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling, et al., in: *Monoclonal Antibodies and T-Cell Hybridomas* 563-681 (Elsevier, N.Y., 1981) (said references incorporated by reference in their entireties). The term "monoclonal antibody" as used herein is not limited to antibodies produced through hybridoma technology. The term "monoclonal antibody" refers to an antibody that is derived from a single clone, including any eukaryotic, prokaryotic, or phage clone, and not the method by which it is produced.

Methods for producing and screening for specific antibodies using hybridoma technology are routine and well known in the art and are discussed in detail in the Examples. In a non-limiting example, mice can be immunized with a polypeptide of the invention or a cell expressing such peptide. Once an immune response is detected, e.g., antibodies specific for the antigen are detected in the mouse serum, the mouse spleen is harvested and splenocytes isolated. The splenocytes are then fused by well known techniques to any suitable myeloma cells, for example cells from cell line SP20 available from the ATCC. Hybridomas are selected and cloned by limited dilution. The hybridoma clones are then assayed by methods known in the art for cells that secrete antibodies capable of binding a polypeptide of the invention. Ascites fluid, which generally contains high levels of antibodies, can be generated by immunizing mice with positive hybridoma clones.

Accordingly, the present invention provides methods of generating monoclonal antibodies as well as antibodies produced by the method comprising culturing a hybridoma cell secreting an antibody of the invention wherein, preferably, the hybridoma is generated by

fusing splenocytes isolated from a mouse immunized with an antigen of the invention with myeloma cells and then screening the hybridomas resulting from the fusion for hybridoma clones that secrete an antibody able to bind a polypeptide of the invention.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, Fab and F(ab')₂ fragments of the invention may be produced by proteolytic cleavage of immunoglobulin molecules, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). F(ab')₂ fragments contain the variable region, the light chain constant region and the CH1 domain of the heavy chain.

For example, the antibodies of the present invention can also be generated using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In a particular embodiment, such phage can be utilized to display antigen binding domains expressed from a repertoire or combinatorial antibody library (e.g., human or murine). Phage expressing an antigen binding domain that binds the antigen of interest can be selected or identified with antigen, e.g., using labeled antigen or antigen bound or captured to a solid surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 binding domains expressed from phage with Fab, Fv or disulfide stabilized Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman et al., J. Immunol. Methods 182:41-50 (1995); Ames et al., J. Immunol. Methods 184:177-186 (1995); Kettleborough et al., Eur. J. Immunol. 24:952-958 (1994); Persic et al., Gene 187 9-18 (1997); Burton et al., Advances in Immunology 57:191-280 (1994); PCT application No. PCT/GB91/01134; PCT publications WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and U.S. Patent Nos. 5,698,426; 5,223,409; 5,403,484; 5,580,717; 5,427,908; 5,750,753; 5,821,047; 5,571,698; 5,427,908; 5,516,637; 5,780,225; 5,658,727; 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

As described in the above references, after phage selection, the antibody coding regions from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host, including mammalian cells, insect cells, plant cells, yeast, and bacteria, e.g., as

described in detail below. For example, techniques to recombinantly produce Fab, Fab' and F(ab')₂ fragments can also be employed using methods known in the art such as those disclosed in PCT publication WO 92/22324; Mullinax et al., *BioTechniques* 12(6):864-869 (1992); and Sawai et al., *AJRI* 34:26-34 (1995); and Better et al., *Science* 240:1041-1043
5 (1988) (said references incorporated by reference in their entirety).

Examples of techniques which can be used to produce single-chain Fvs and antibodies include those described in U.S. Patents 4,946,778 and 5,258,498; Huston et al., *Methods in Enzymology* 203:46-88 (1991); Shu et al., *PNAS* 90:7995-7999 (1993); and Skerra et al., *Science* 240:1038-1040 (1988). For some uses, including in vivo use of antibodies in
10 humans and in vitro detection assays, it may be preferable to use chimeric, humanized, or human antibodies. A chimeric antibody is a molecule in which different portions of the antibody are derived from different animal species, such as antibodies having a variable region derived from a murine monoclonal antibody and a human immunoglobulin constant region. Methods for producing chimeric antibodies are known in the art. See e.g., Morrison,
15 *Science* 229:1202 (1985); Oi et al., *BioTechniques* 4:214 (1986); Gillies et al., (1989) *J. Immunol. Methods* 125:191-202; U.S. Patent Nos. 5,807,715; 4,816,567; and 4,816,397, which are incorporated herein by reference in their entirety. Humanized antibodies are antibody molecules from non-human species antibody that binds the desired antigen having one or more complementarity determining regions (CDRs) from the non-human species and
20 a framework regions from a human immunoglobulin molecule. Often, framework residues in the human framework regions will be substituted with the corresponding residue from the CDR donor antibody to alter, preferably improve, antigen binding. These framework substitutions are identified by methods well known in the art, e.g., by modeling of the interactions of the CDR and framework residues to identify framework residues important
25 for antigen binding and sequence comparison to identify unusual framework residues at particular positions. (See, e.g., Queen et al., U.S. Patent No. 5,585,089; Riechmann et al., *Nature* 332:323 (1988), which are incorporated herein by reference in their entirety.) Antibodies can be humanized using a variety of techniques known in the art including, for example, CDR-grafting (EP 239,400; PCT publication WO 91/09967; U.S. Patent Nos.
30 5,225,539; 5,530,101; and 5,585,089), veneering or resurfacing (EP 592,106; EP 519,596; Padlan, *Molecular Immunology* 28(4/5):489-498 (1991); Studnicka et al., *Protein*

Engineering 7(6):805-814 (1994); Roguska. et al., PNAS 91:969-973 (1994)), and chain shuffling (U.S. Patent No. 5,565,332).

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Human antibodies can be made by a variety of methods known in the art including phage display methods described above using antibody libraries derived from human immunoglobulin sequences. See also, U.S. Patent Nos. 4,444,887 and 4,716,111; and PCT publications WO 98/46645, WO 98/50433, WO 98/24893, WO 98/16654, WO 96/34096, WO 96/33735, and WO 91/10741; each of which is incorporated herein by reference in its entirety.

Human antibodies can also be produced using transgenic mice which are incapable of expressing functional endogenous immunoglobulins, but which can express human immunoglobulin genes. For example, the human heavy and light chain immunoglobulin gene complexes may be introduced randomly or by homologous recombination into mouse embryonic stem cells. Alternatively, the human variable region, constant region, and diversity region may be introduced into mouse embryonic stem cells in addition to the human heavy and light chain genes. The mouse heavy and light chain immunoglobulin genes may be rendered non-functional separately or simultaneously with the introduction of human immunoglobulin loci by homologous recombination. In particular, homozygous deletion of the JH region prevents endogenous antibody production. The modified embryonic stem cells are expanded and microinjected into blastocysts to produce chimeric mice. The chimeric mice are then bred to produce homozygous offspring which express human antibodies. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained from the immunized, transgenic mice using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA, IgM and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar, Int. Rev. Immunol. 13:65-93 (1995). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., PCT publications WO 98/24893; WO 92/01047; WO 96/34096; WO 96/33735; European Patent

No. 0 598 877; U.S. Patent Nos. 5,413,923; 5,625,126; 5,633,425; 5,569,825; 5,661,016; 5,545,806; 5,814,318; 5,885,793; 5,916,771; and 5,939,598, which are incorporated by reference herein in their entirety. In addition, companies such as Abgenix, Inc. (Freemont, CA) and Genpharm (San Jose, CA) can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a mouse antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al., Bio/technology 12:899-903 (1988)).

Further, antibodies to the polypeptides of the invention can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" polypeptides of the invention using techniques well known to those skilled in the art. (See, e.g., Greenspan & Bona, FASEB J. 7(5):437-444; (1989) and Nissinoff, J. Immunol. 147(8):2429-2438 (1991)). For example, antibodies which bind to and competitively inhibit polypeptide multimerization and/or binding of a polypeptide of the invention to a ligand can be used to generate anti-idiotypes that "mimic" the polypeptide multimerization and/or binding domain and, as a consequence, bind to and neutralize polypeptide and/or its ligand. Such neutralizing anti-idiotypes or Fab fragments of such anti-idiotypes can be used in therapeutic regimens to neutralize polypeptide ligand. For example, such anti-idiotypic antibodies can be used to bind a polypeptide of the invention and/or to bind its ligands/receptors, and thereby block its biological activity.

Polynucleotides Encoding Antibodies

The invention further provides polynucleotides comprising a nucleotide sequence encoding an antibody of the invention and fragments thereof. The invention also encompasses polynucleotides that hybridize under stringent or alternatively, under lower stringency hybridization conditions, e.g., as defined supra, to polynucleotides that encode an antibody, preferably, that specifically binds to a polypeptide of the invention, preferably, an antibody that binds to a polypeptide having the amino acid sequence of SEQ ID NO:Y.

The polynucleotides may be obtained, and the nucleotide sequence of the polynucleotides determined, by any method known in the art. For example, if the nucleotide sequence of the antibody is known, a polynucleotide encoding the antibody may be

assembled from chemically synthesized oligonucleotides (e.g., as described in Kutmeier et al., *BioTechniques* 17:242 (1994)), which, briefly, involves the synthesis of overlapping oligonucleotides containing portions of the sequence encoding the antibody, annealing and ligating of those oligonucleotides, and then amplification of the ligated oligonucleotides by PCR.

Alternatively, a polynucleotide encoding an antibody may be generated from nucleic acid from a suitable source. If a clone containing a nucleic acid encoding a particular antibody is not available, but the sequence of the antibody molecule is known, a nucleic acid encoding the immunoglobulin may be chemically synthesized or obtained from a suitable source (e.g., an antibody cDNA library, or a cDNA library generated from, or nucleic acid, preferably poly A⁺ RNA, isolated from, any tissue or cells expressing the antibody, such as hybridoma cells selected to express an antibody of the invention) by PCR amplification using synthetic primers hybridizable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe specific for the particular gene sequence to identify, e.g., a cDNA clone from a cDNA library that encodes the antibody. Amplified nucleic acids generated by PCR may then be cloned into replicable cloning vectors using any method well known in the art.

Once the nucleotide sequence and corresponding amino acid sequence of the antibody is determined, the nucleotide sequence of the antibody may be manipulated using methods well known in the art for the manipulation of nucleotide sequences, e.g., recombinant DNA techniques, site directed mutagenesis, PCR, etc. (see, for example, the techniques described in Sambrook et al., 1990, *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and Ausubel et al., eds., 1998, *Current Protocols in Molecular Biology*, John Wiley & Sons, NY, which are both incorporated by reference herein in their entireties), to generate antibodies having a different amino acid sequence, for example to create amino acid substitutions, deletions, and/or insertions.

In a specific embodiment, the amino acid sequence of the heavy and/or light chain variable domains may be inspected to identify the sequences of the complementarity determining regions (CDRs) by methods that are well known in the art, e.g., by comparison to known amino acid sequences of other heavy and light chain variable regions to determine the regions of sequence hypervariability. Using routine recombinant DNA techniques, one or more of the CDRs may be inserted within framework regions, e.g., into human framework

regions to humanize a non-human antibody, as described supra. The framework regions may be naturally occurring or consensus framework regions, and preferably human framework regions (see, e.g., Chothia et al., J. Mol. Biol. 278: 457-479 (1998) for a listing of human framework regions). Preferably, the polynucleotide generated by the combination of the
5 framework regions and CDRs encodes an antibody that specifically binds a polypeptide of the invention. Preferably, as discussed supra, one or more amino acid substitutions may be made within the framework regions, and, preferably, the amino acid substitutions improve binding of the antibody to its antigen. Additionally, such methods may be used to make amino acid substitutions or deletions of one or more variable region cysteine residues
10 participating in an intrachain disulfide bond to generate antibody molecules lacking one or more intrachain disulfide bonds. Other alterations to the polynucleotide are encompassed by the present invention and within the skill of the art.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., Proc. Natl. Acad. Sci. 81:851-855 (1984); Neuberger et al., Nature
15 312:604-608 (1984); Takeda et al., Nature 314:452-454 (1985)) by splicing genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. As described supra, a chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a
20 human immunoglobulin constant region, e.g., humanized antibodies.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778; Bird, Science 242:423-42 (1988); Huston et al., Proc. Natl. Acad. Sci. USA 85:5879-5883 (1988); and Ward et al., Nature 334:544-54 (1989)) can be adapted to produce single chain antibodies. Single chain antibodies are formed by linking the heavy
25 and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide. Techniques for the assembly of functional Fv fragments in E. coli may also be used (Skerra et al., Science 242:1038-1041 (1988)).

Methods of Producing Antibodies

30 The antibodies of the invention can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or preferably, by recombinant expression techniques.

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
 - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X;
 - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X;
 - (c) a polynucleotide encoding a polypeptide fragment of a polypeptide encoded by SEQ ID NO:X or a polypeptide fragment encoded by the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X;
 - (d) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X;
 - (e) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X;
 - (f) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X, having biological activity;
 - (g) a polynucleotide which is a variant of SEQ ID NO:X;
 - (h) a polynucleotide which is an allelic variant of SEQ ID NO:X;
 - (i) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
 - (j) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(i), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide

sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a protein.

5

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X.

10

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in the related cDNA clone, which is hybridizable to SEQ ID NO:X.

15

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

20

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

25

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

30

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least
5 95% identical to a sequence selected from the group consisting of:
- (a) a polypeptide fragment of SEQ ID NO:Y or of the sequence encoded by the cDNA included in the related cDNA clone;
 - (b) a polypeptide fragment of SEQ ID NO:Y or of the sequence encoded by the cDNA included in the related cDNA clone, having biological activity;
 - 10 (c) a polypeptide domain of SEQ ID NO:Y or of the sequence encoded by the cDNA included in the related cDNA clone;
 - (d) a polypeptide epitope of SEQ ID NO:Y or of the sequence encoded by the cDNA included in the related cDNA clone;
 - (e) a full length protein of SEQ ID NO:Y or of the sequence encoded by the
15 cDNA included in the related cDNA clone;
 - (f) a variant of SEQ ID NO:Y;
 - (g) an allelic variant of SEQ ID NO:Y; or
 - (h) a species homologue of the SEQ ID NO:Y.
12. The isolated polypeptide of claim 11, wherein the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.
13. An isolated antibody that binds specifically to the isolated polypeptide
25 of claim 11.
14. A recombinant host cell that expresses the isolated polypeptide of claim 11.
15. A method of making an isolated polypeptide comprising:
- 30

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

5 16. The polypeptide produced by claim 15.

 17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

10

 18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

 (a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

15 (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

 19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

20 (a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

 (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

25 20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

 (a) contacting the polypeptide of claim 11 with a binding partner; and

 (b) determining whether the binding partner effects an activity of the polypeptide.

30

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.
22. A method of identifying an activity in a biological assay, wherein the method comprises:
- 5 (a) expressing SEQ ID NO:X in a cell;
- (b) isolating the supernatant;
- (c) detecting an activity in a biological assay; and
- (d) identifying the protein in the supernatant having the activity.
- 10 23. The product produced by the method of claim 20.

<212> DNA

<213> Homo sapiens

<400> 218

```
ggcccggtgg ggccggccgg gactcgccgc tcgcacgccc ttgggcccgg gccgggccc 60
cgctcttctt tccgcttgcg ctgtgagctg aggcggtgta tgtgcggcaa taacatgtca 120
accccgctgc ccgccatcgt gcccgcgcgc cggaaggcca ccgctgcggt gattttcctg 180
catggattgg gagatactgg gcacggatgg gcagaagcct ttgcaggatc cagaagttca 240
catatcaaat atatctgccc gcatgcgcct gttaggcctg ttacattaaa tatgaacgtg 300
gctatgcctt catggtttga tattattggg ctttcaccag attcacagga ggatgaatct 360
gggattaaac aggcagcaga aaatataaaa gctttgattg atcaagaagt gaagaatggc 420
attccttcta acagaattat tttgggaggg ttttctcagg gaggagcttt atctttatat 480
actgccttta ccacacagca gaaactggca ggtgtcactg cactcagttg ctggcttcca 540
cttcgggctt cctttccaca gggtcctatc ggtggtgcta atagagatat ttctattctc 600
cagtgccacg gggattgtga ccctttggtt cccctgatgt ttggttctct tacgggtgaa 660
aaactaaaaa cattggtgaa tccagccaat gtgaccttta aaacctatga aggtatgatg 720
cacagttcgt gtcaacagga aatgatggat gtcaagcaat tcattgataa actcctacct 780
ccaattgatt gacgtcacta agaggccttg tgtagaagta caccagcatc attgtagtag 840
agtgtaaacc ttttcccatg cccagtcttc aaatttctaa tgttttgcag tgttaaaatg 900
ttttgcaaat acatgccaat aacacagatc aaataatata tcctcatgag aaatttatga 960
tcttttaagt ttctatacat gtattcttat aagacgaccc aggatctact atattagaat 1020
agatgaagca ggtagcttct tttttctcaa atgtaattca gcaaaataat acagtactgc 1080
caccagattt ttattacat catttgaaaa ttagcagtat gcttaatgaa aatttgttca 1140
ggtataaatg agcagttaag atataaacia tttatgcatg ctgtgactta gtctatggat 1200
ttattccaaa attgcttagt caccatgcag tgtctgtatt tttatatatg tgttcatata 1260
tacataatga ttataatata taataagaat gaggtggtat tacattattc ctaataatag 1320
ggataatgct gtttattgtc aagaaaaagt aaaatcggtc tcttcaatta atggcccttt 1380
tattttggga ccaggctttt attttccctg atattatttc tatttaatac tcttttctct 1440
caagaaaaaa aaaaaagttt gtttttctt tattgtcctt catagcaggc caagtattgc 1500
ctctctgcaa tagacagcta ctgtcaatac atgctgtaat ttgacattct gggtcacaga 1560
tataaggtat ttaaaatcta tttatgcttt atagagaaac cagacattaa aacttcatgc 1620
actacttatt tcgaattact gtaccttacc caaatttaca cctagctatt aggatcttca 1680
accaggttaa caggaataat tctgtggttt catttttctg taaacaactg aaagaataat 1740
tagatcatat tctagtatgt tctgaaatat ctttaagact gatcttaaaa actaacttct 1800
aagatgattt catcttctca tagtatagag tttactttgt acacgtttga aaccaactac 1860
tgtagaagat gaggaatcta ttgtaatttt ttgctttatt ttcattctgcc agtggactta 1920
tttgaaattt tcactttagt caaattattt tttgtattag tttttgatgc agacataaaa 1980
atagcaatca ttttaaattg tcaaaatttc cagattactg gtaaaaatta tttgaaaaca 2040
aacttatggg taataaaggc tagtcagaac cctataccat aaagtgtagt taccatacag 2100
attaatatgt agcaaaaatg tatgcttgat atttctcaac tgtgttaatt tttctgctgt 2160
attccagctg accaaaacaa tattaagaat gcattcttat aaatgggtgc taattgataa 2220
tggaataaat ttagtaatgg actatacagg atgttaataa tgaagccata tgtttatgtc 2280
tggatttaaa aattttaaac aatcatttac tatgtcattt ttctttacct tgaagaacat 2340
aaactgttat ttcacttcta caaatcagca agatattatt tatggcaaga aatattccat 2400
tgaaaatatt tgctgtaaca tgggaaagtg taaatgtttt tcatggtttc tatcaatgtg 2460
aaataaaaatt taattctgaa aaaaaaaaaa aaa 2493
```

<210> 219

<211> 1259

<212> DNA

<213> Homo sapiens

<400> 219

gcgcccgcggt gccggaaccg ctgggcggga gcgaggcggg gcggctgcag ctgcagggcg 60
aggagctgcg gctgcaggag gagagcgtgc ggctgcacca gattaacatc tacctcagcg 120
accgcatctc actgcaccgc cgcctgcccg wgcgctggaa cccgctgtgc aaagagaaga 180
aatatgatta tgataatttg cccaggacat ctgttatcat agcattttat aatgaagcct 240
ggtcaactct ccttcggaca gtttacagtgc tccttgagac atccccggat atcctgctag 300
aagaagtgat ccttgtagat gactacagtgc atagagagca cctgaaggag cgcttgccca 360
atgagctttc gggactgccc aagggtgcgc tgatccgcgc caacaagaga gagggcctgg 420
tgcgagcccg gctgctgggg gcgtctgcgg cgargggcga tgttctgacc ttcctggact 480
gtcactgtga gtgccacgaa ggggtgctgga gccgctgctg cagaggatcc atgaagagga 540
gtcggcagtg gtgtgcccg tgattgatgt satcgactgg aacaccttcg aatacctggg 600
gaactccggg gagccccaga tcggcggttt cgactggagg ctgggtgtca cgtggcacac 660
agttcctgag agggagagga tacggatgca atccccgctc gatgtcatca ggtctccaac 720
aatggctggg gggctgtttg ctgtgagtaa gaaatatttt gaatatctgg ggtcttatga 780
tacaggaatg gaagtttggg gaggagaaaa cctcgaattt tccttttaga tctggcagtg 840
tggtggggtt ctggaacac acccatgttc ccatgttggc catgttttcc ccaagcaagc 900
tccctactcc cgcaacaagg ctctggccaa cagtgttcgt gcagctgaag tatggatgga 960
tgaatttaaa gagctctact accatcgcaa cccccgtgcc cgcttggaac cttttgggga 1020
tgtgacagag aggaagcagc tccgggacaa gctccagtgt aaagacttca agtggttctt 1080
ggagactgtg tatccagaac tgcattgtgc tgaggacagg cctggcttct tcgggatgct 1140
ccagaacaaa ggactaacag actactgctt tgactwtaac cctcccgatg aaaaccagat 1200
tgtgggacac caggtcattc tgtacctctg tcatgggatg ggccagaatg acctggtgc 1259

<210> 220

<211> 1849

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (920)

<223> n equals a,t,g, or c

<400> 220

ctgttgatg gagcaggggt tgtgggtttt ctgggcccac cattatggct gcttcagagt 60
cagaagaaag ccatagggca gtaggggagc tcctattgcc tagccccctc ccctttgtgg 120
ctcccactct agctgcctat ttttgctcat cagctgggtga gtcagtatgg gccagcagtt 180
ctccctccct aagcccttgc tactttatgg gtttagctttg caggtttggg ggcttgaggg 240
gtgggggcaa ctcaccactg ccaggtaact ccctgaaggg tgggagtggg ttatcttcta 300
ggctcttacc cgcggtaggg aagggcacac aacctgtctt ccttccattc tcctttcccc 360
catcccattt agtgctgcca cagggcagaa gcacacaaac caaccacaca gtctctgact 420
tctcctaagc actttgagtt gttgaatggg gctcaggggc aagagttttt gctgccctcc 480
ccagcgtggg cacagggtta ttgaactgcc tgcacttggt tctcatgcaa ctccagcatt 540
ttcccagaa gttgaactat ggatagcagc ttggtatgga tttcctaaat cttaacattt 600
gaagcagctt cttgaggctg gcaactatcc tggtttctgt cttggagggg gtggtttgtt 660
tgctggggcc caacgtctgt cccaagtggg ggggtgagag taagttaact ttggtgccag 720
gtgagaggtg ggggctcttt gcttagactc cctatcatgg aaagattgga gttttctatg 780
cagggcactg gggaaaagga ttgctgattc tgactgacct tgatcagaga gattaggatt 840
gtattttgac ataggatttg gaacctatct aaatgttgaa gttccctgag acagctctcc 900
agctgctgag cctgcgccan gggctaagca gccccctaat agaggctctg ctccctttcc 960

His Ile Lys Tyr Ile Cys Pro His Ala Pro Val Arg Pro Val Thr Leu
 85 90 95
 Asn Met Asn Val Ala Met Pro Ser Trp Phe Asp Ile Ile Gly Leu Ser
 100 105 110
 Pro Asp Ser Gln Glu Asp Glu Ser Gly Ile Lys Gln Ala Ala Glu Asn
 115 120 125
 Ile Lys Ala Leu Ile Asp Gln Glu Val Lys Asn Gly Ile Pro Ser Asn
 130 135 140
 Arg Ile Ile Leu Gly Gly Phe Ser Gln Gly Gly Ala Leu Ser Leu Tyr
 145 150 155 160
 Thr Ala Leu Thr Thr Gln Gln Lys Leu Ala Gly Val Thr Ala Leu Ser
 165 170 175
 Cys Trp Leu Pro Leu Arg Ala Ser Phe Pro Gln Gly Pro Ile Gly Gly
 180 185 190
 Ala Asn Arg Asp Ile Ser Ile Leu Gln Cys His Gly Asp Cys Asp Pro
 195 200 205
 Leu Val Pro Leu Met Phe Gly Ser Leu Thr Val Glu Lys Leu Lys Thr
 210 215 220
 Leu Val Asn Pro Ala Asn Val Thr Phe Lys Thr Tyr Glu Gly Met Met
 225 230 235 240
 His Ser Ser Cys Gln Gln Glu Met Met Asp Val Lys Gln Phe Ile Asp
 245 250 255
 Lys Leu Leu Pro Pro Ile Asp
 260

<210> 992

<211> 256

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (229)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 992

Val Pro Arg Arg Val Leu Glu Pro Leu Leu Gln Arg Ile His Glu Glu

1	5	10	15
Glu Ser Ala Val Val Cys Pro Val Ile Asp Val Ile Asp Trp Asn Thr	20	25	30
Phe Glu Tyr Leu Gly Asn Ser Gly Glu Pro Gln Ile Gly Gly Phe Asp	35	40	45
Trp Arg Leu Val Phe Thr Trp His Thr Val Pro Glu Arg Glu Arg Ile	50	55	60
Arg Met Gln Ser Pro Val Asp Val Ile Arg Ser Pro Thr Met Ala Gly	65	70	75
Gly Leu Phe Ala Val Ser Lys Lys Tyr Phe Glu Tyr Leu Gly Ser Tyr	85	90	95
Asp Thr Gly Met Glu Val Trp Gly Gly Glu Asn Leu Glu Phe Ser Phe	100	105	110
Arg Ile Trp Gln Cys Gly Gly Val Leu Glu Thr His Pro Cys Ser His	115	120	125
Val Gly His Val Phe Pro Lys Gln Ala Pro Tyr Ser Arg Asn Lys Ala	130	135	140
Leu Ala Asn Ser Val Arg Ala Ala Glu Val Trp Met Asp Glu Phe Lys	145	150	155
Glu Leu Tyr Tyr His Arg Asn Pro Arg Ala Arg Leu Glu Pro Phe Gly	165	170	175
Asp Val Thr Glu Arg Lys Gln Leu Arg Asp Lys Leu Gln Cys Lys Asp	180	185	190
Phe Lys Trp Phe Leu Glu Thr Val Tyr Pro Glu Leu His Val Pro Glu	195	200	205
Asp Arg Pro Gly Phe Phe Gly Met Leu Gln Asn Lys Gly Leu Thr Asp	210	215	220
Tyr Cys Phe Asp Xaa Asn Pro Pro Asp Glu Asn Gln Ile Val Gly His	225	230	235
Gln Val Ile Leu Tyr Leu Cys His Gly Met Gly Gln Asn Asp Leu Val	245	250	255

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/05883

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12P 21/04; C12N 15/00; C07H 21/02
US CL : 435/70.1, 320.1; 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/70.1, 320.1; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---	SCANLAN et al. Characterization of Human Colon Cancer Antigens Recognized by Autologous Antibodies, Int. J. Cancer, 1998, Vol. 76, pages 652-658.	1-4, 11-12, 16 -----
Y		5-10, 14-15
X ---	TANAKA et al. A Novel Variant of Human Grb7 Is Associated with Invasive Esophageal Carcinoma, J. Clin. Invest., August 1998, Vol. 102, No. 4, pages 821-827.	1-4, 11-12, 16 -----
Y		5-10, 14-15
X ---	KISHI et al. Molecular Cloning of Human GRB-7 Co-amplified with CAB1 and c-ERBB-2 in Primary Gastric Cancer, Biochemical and Biophysical Research Communications, 1997, Vol. 232, pages 5-9.	1-4, 11-12, 16 -----
Y		5-10, 14-15
X ---	JIANG et al. Subtraction hybridization identifies a novel melanoma differentiation associated gene, mda-7, modulated during human melanoma differentiation, growth and progression, Oncogenes, 1995, Vol. 11, pages 2477-2486.	1-4, 11-12, 16 -----
Y		5-10, 14-15
X ---	MUELLER et al. Polymerase Chain Reaction Selects a Novel Disintegrin Proteinase from CD40-Activated Germinal Center Dendritic Cells, J. Exp. Med., August 1997, Vol. 186, No. 5, pages 655-663.	1-4, 11-12, 16 -----
Y		5-10, 14-15

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

18 May 2000 (18.05.2000)

Date of mailing of the international search report

13 JUN 2000

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Young J. Kim

Telephone No. (703) 308-0196

Jerry J. Deyfor

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/05883

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	FOJO et al. Donor Splice Site Mutation in the Apolipoprotein (Apo) C-II Gene (APO C-IIhamburg) of a Patient with APO C-II Deficiency, The Journal of Clinical Investigations, November 1988, Vol. 82, pages 1489-1494.	1-4, 11-12, 16 ----- 5-10, 14-15
X --- Y	JACKSON et al. Isolation of cDNA and Genomic Clones for Apolipoprotein C-II, Methods in Enzymology, 1986, Vol. 128, pages 788-800.	1-4, 11-12, 16 ----- 5-10, 14-15
X --- Y	HILLIER et al. Generation and Analysis of 280,000 Human Expressed Sequence Tags, Genome Research, 1996, Vol. 6, No. 9, pages 807-828.	1-4, 11-12, 16 ----- 5-10, 14-15
Y	WATSON et al. The Science Used in the Recombinant DNA Industry. In: Recombinant DNA, W.H. Freeman and Company, 1983, pages 231-241.	7-10, 14-15

INTERNATIONAL SEARCH REPORT

Intern. application No.

PCT/US00/05883

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-12,14,15,16,21

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/05883

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-12, 14, 15, 16, and 21, drawn to cDNA, polypeptides, genes, a method of using the cDNA to make host cells comprising the cDNA, and a method of making the polypeptide.

Group II, claim(s) 13, drawn to an antibody specific for the polypeptides of Group I.

Group III, claim(s) 17, drawn to a therapeutic method of using the cDNA or the polypeptide of Group I.

Group IV, claim(s) 18 and 19, drawn to a diagnostic method of using the cDNA or polypeptide of Group I.

Group V, claim(s) 20, drawn to a method of using the polypeptide of Group I to isolate a binding partner.

Group VI, claim(s) 22, drawn to a method of using the cDNA of Group I to identify the activity of the polypeptide encoded by the cDNA.

Group VII, claim(s) 23, drawn to the binding partner made by the method of Group V.

The inventions listed as Groups I-VII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: PCT Rule 13.1 and Annex B do not provide for unity of invention between two or more different products or methods of use that share a special technical feature.

In addition, each Group detailed above reads on distinct Groups drawn to multiple SEQ ID Numbers. The sequences are distinct because they are unrelated sequences, and a further lack of unity is applied to each Group. The lack of unity is partially waived and the Applicants must further elect 10 SEQ ID Numbers for examination in the elected Group detailed above.

Continuation of B. FIELDS SEARCHED Item 3: SEQUENCE DATABASES (US PATENT, INTERFERENCE, COMMERCIAL)

STN COMMERCIAL DATABASE (Biosis, Medline, Embase, Embal, SciSearch, BiotechDS, CaPlus)

Search Terms: Recombinant, Host, Cell, Vector, peptide, protein, cDNA